

Assessment and translation of novel biomarkers for diagnosis of maturity onset diabetes of the young due to *HNF1A* variants

A thesis submitted to the University of Oxford



In candidature for the degree of Doctor of Philosophy

by

Agata Juszczak

Harris Manchester College, University of Oxford

and

Oxford Centre for Diabetes, Endocrinology and Metabolism

Radcliffe Department of Medicine, University of Oxford

Hilary term, 2018

Memorandum

The work presented in this thesis is the original work of the author. DNA sequencing and functional studies have been performed by the author in the laboratories of the Oxford Centre for Diabetes, Endocrinology and Metabolism under the supervision of Professor Mark McCarthy, Professor Anna Gloyn and Associate Professor Katherine Owen. Funding was provided by Diabetes UK George Alberti Research Fellowship, the European Foundation for the Study of Diabetes (EFSD) and the NIHR Oxford Biomedical Research Centre.

The measurement of plasma N-glycan profile, described in chapter 5, was performed by Ms Tamara Pavic in the Laboratories of the Faculty of Pharmacy and Biochemistry, University of Zagreb under the supervision of Professor Gordan Lauc and Dr Olga Gornik. However, an assessment of glycans as potential biomarkers would not be possible without *HNF1A* sequencing and interpretation of the variants which I have performed. I have analysed the results and described them in chapter 5. This work is included in the manuscript under joint first authorship currently being finalized.

The measurement of the urinary steroid metabolites, included in chapter 6, was performed in the Laboratory of the Institute of Metabolism and Systems Research, University of Birmingham by Beverly Hughes and Lorna Gilligan under supervision of Professor Wiebke Arlt. I have analysed the results and I have written a manuscript as a first author which is under revision of supervisors at current stage.

I hereby state that this thesis has not been submitted for any other degree at this or other university.

This thesis is approximately 48,000 words.

Assessment and translation of novel biomarkers for diagnosis of maturity onset diabetes
of the young due to *HNF1A* variants

Agata Juszcak, Harris Manchester College

Submitted in candidature for the degree of Doctor of Philosophy, Hilary Term 2018

HNF1A-MODY is the most common form of monogenic diabetes, but due to low awareness among clinicians and limited availability of genetic testing in many countries, the majority of cases are misdiagnosed as type 1 or type 2 diabetes. In clinical practice, it is apparent that some young adults with non-autoimmune diabetes share several features of MODY and type 2 diabetes. Hence, I aimed to understand the range of genetic variation in *HNF1A* in this group and whether we could use biomarkers to help identify those with *HNF1A* alleles likely to be influencing their diabetes status.

I performed *HNF1A* sequencing in 1058 subjects with non-autoimmune diabetes diagnosed below 45 years of age and identified 26 rare non-synonymous *HNF1A* alleles in 30 probands and 8 relatives. A systematic assessment assigned a likely damaging functional effect to 50% of these *HNF1A* alleles. The assessment of *HNF1A* alleles and their experimental characterisation unveiled the complexity of the genotype-phenotype relationship and continuity of the functional spectrum of *HNF1A* alleles from benign through likely benign, likely damaging to damaging effect.

To clarify the performance of C-reactive protein (CRP) and N-glycan profile as biomarkers of HNF1A-MODY, I evaluated them in the above unselected sample of individuals who also underwent *HNF1A* sequencing. I have shown that 3-antennary fucosylated N-glycans and CRP were significantly lower in subjects with likely damaging *HNF1A* alleles in this group when compared to participants without *HNF1A* variants. Both biomarkers distinguished individuals with likely damaging *HNF1A* alleles from those without *HNF1A* variants with a C-statistic of 0.82-0.92. Plasma levels of these biomarkers negatively correlated with the predicted pathogenicity of *HNF1A* alleles.

The expression of several steroidogenesis enzymes had been reported to be reduced in the liver cells of *Hnf1a* knock-out mice. Hence, I assessed the urinary steroid profile in individuals with HNF1A-MODY and age- and BMI-matched non-diabetic subjects and those with type 2 diabetes. The activity of 11 β -hydroxysteroid dehydrogenase type 2 was reduced and the activity of 5 β -steroid reductase increased in individuals with HNF1A-MODY. Their performance in selecting subjects with HNF1A-MODY was inferior to that of CRP and 3-antennary fucosylated glycans.

In summary, my thesis has shown the complexity of the interpretation of genetic variation which is crucial for an accurate diagnosis and the assessment of potential biomarkers for genetic disease. Careful evaluation of the functional effect of *HNF1A* alleles identified on sequencing informs diagnosis in individuals with early-onset diabetes. Furthermore, the combination of functional assessment and clinical phenotype can guide management in those without canonical monogenic phenotypes found to have variant *HNF1A* alleles.

Acknowledgements

I would like to thank my supervisors Professor Mark McCarthy and Professor Katharine Owen for the opportunity, resources and guidance to complete my thesis. I am very grateful to Professor Mark McCarthy for setting an exceptional example and sharing his knowledge on genetics and genomics of diabetes. Thank you to Professor Katharine Owen for introducing me to the world of monogenic diabetes and giving me an opportunity to participate in the MODY clinic to learn the management of the most common forms of monogenic diabetes.

I would like to thank Professor Anna Gloyn for letting me use her laboratory and supervising my functional studies, and to Prof Jeremy Tomlinson for guiding me through the maze of steroid metabolites.

A special thank you to Dr Amanda Bennett, Chris Groves and Carla Burrows for introducing me to the laboratory work, teaching me the essential laboratory skills and answering multiple scientific questions crucial for my progress. Also, thanks to Nicola Beer and Anne Raimondo for the useful discussions related to my experiments.

Finally, I want to thank my husband, Maciek, for his patience and support, and my daughter, Maria for reading every chapter.

List of common abbreviations

1,5-AG	1,5-anhydroglucitol
17HP	17-hydroxy-prognanolone
5ASRD	5 α -steroid reductase
5ATHF	5-alpha-tetrahydrocortisone
5BSRD	5 β -steroid reductase
5BTHF	5-beta-tetrahydrocortisone
5PD	pregnenediol
5PT	5-pregnenetriol
AA	amino acid
ABCC8	ATP binding cassette subfamily C member 8
ACTH	adrenocorticotrophic hormone
AKR1C1-4	aldo-keto-reductase family 1 member C1-4
AKR1D1	aldo-keto-reductase family 1 member D1
AUC	area under the curve
B	corticosterone
BMI	body mass index
CEL	carboxyl ester lipase
CYP7A1	cytochrome P450 family 7 subfamily A member 1
CYP7B1	cytochrome P450 family 7 subfamily B member 1
DG	desialylated glycan
DHEA	dehydroepiandrosterone
DHT	dihydrotestosterone
DOC	11-desoxycorticosterone
E	cortisone
ESP	Exome Sequencing Project
ExAC	Exome Aggregation Consortium
F	cortisol
GAD	glutamic acid decarboxylase
GC	glucocorticoids

GC/MS	gas chromatography/ mass spectrometry
GCK	glucokinase
GDP-fucose	guanosine-diphosphate-fucose
GLUT2	SLC2A2, solute carrier family 2 member 2
GnomAD	Genome Aggregation Database
GP	glycoprotein
GWAS	genome wide association study
HGMD	Human Gene Variation Database
HGMD	Human Gene Mutation Database
HILIC-UPLC	Hydrophilic Interaction Chromatography Ultra-performance liquid chromatography
HNF1A	hepatocyte nuclear factor 1-alpha
HNF1B	hepatocyte nuclear factor 4-alpha
HNF4A	hepatocyte nuclear factor 1-beta
HPLC	high performance liquid chromatography
hsCRP	high-sensitivity C-reactive protein
HSD11B1	11-beta-hydroxysteroid dehydrogenase type 1
HSD11B2	11-beta-hydroxysteroid dehydrogenase type 2
HSD20A	20-beta-hydroxysteroid dehydrogenase
HSD20B	20-alpha--hydroxysteroid dehydrogenase
HSD3A	3-alpha-hydroxysteroid dehydrogenase
HSD3B	3-beta-hydroxysteroid dehydrogenase
HSF	Human Splicing Finder
INS	insulin
IQR	interquartile range
IRS-1	insulin receptor substrate-1
KCNJ11	potassium voltage-gated channel subfamily J member 11
LH	luteinising hormone
MODY	Maturity Onset Diabetes of the Young
NEUROD1	neurogenic differentiation 1

NLR	negative likelihood ratio
NPV	negative predictive value
OR	odds ratio
PD	pregnanediol
PDX1	IPF1, insulin promoter factor 1
PEPCK	phosphoenolopyruvate-carboxykinase
PI3K	phosphatidylinositol 3-kinase
PKLR	pyruvate kinase
pLI	the probability of being Loss-of-function Intolerant
PLR	positive likelihood ratio
POU	protein homeodomain
PPV	positive predictive value
PT	pregnanetriol
PTV	protein truncating variant
ROC	receiver operating characteristics
SD	standard deviation
SEARCH	Search for Diabetes in Youth study
SGLT2	SLC5A2, solute carrier family 5 member 2
SN	sensitivity
SNP	single nucleotide polymorphism
SP	specificity
T1D	type 1 diabetes
T2D	type 2 diabetes
T2D-GENES	Type 2 Diabetes Genes study
TCF1	transcription factor 1, HNF1A
THA	tetrahydro-11-dehydrocorticosterone
THB	tetrahydrocorticosterone
THE	tetrahydrocortisone
THF	tetrahydrocortisol
THS	tetrahydro-11-desoxycortisol

UCPCR	urine C-peptide to creatinine ratio
VUS	variant of unknown significance
WHO	World Health Organisation
WT	wild type

Table of Contents

Memorandum	I
Acknowledgements	III
List of common abbreviations.....	IV
Chapter 1.....	1
General introduction	1
1.1 Young adult onset diabetes.....	2
1.2 Monogenic diabetes	3
1.2.1 HNF1A-MODY	5
1.2.2 HNF4A-MODY	10
1.2.3 HNF1B-MODY.....	12
1.2.4 GCK-MODY	13
1.2.5 Maternally inherited diabetes and deafness (MIDD)	15
1.2.6 Other rare forms of MODY	16
1.3 Differential diagnosis of young adult onset diabetes	17
1.3.1 Current challenges in diagnosing MODY	18
1.3.2 Individuals initially diagnosed with Type 1 Diabetes	19
1.3.3 Individuals with initial diagnosis of type 2 diabetes	21
1.4 Hepatocyte nuclear factor 1 alpha (HNF1A)	22
1.5 Challenges in interpretation of genetic variants.....	26
1.6 Biomarkers in HNF1A-MODY	29
1.6.1 Previously investigated biomarkers of HNF1A-MODY	29
1.6.2 Biomarkers of HNF1A-MODY with a translational potential	34
1.7 Adrenal steroids and <i>Hnf1a</i> knock-out mouse	38
1.7.1 Expression of steroidogenesis enzymes in <i>Hnf1a</i> knock-out mice	38
1.7.2 The role of glucocorticoids in glucose metabolism	39
1.7.3 Steroidogenesis pathway	41

1.7.4 Urinary steroid metabolites in human	44
1.8 Aims of the thesis	47
Chapter 2.....	49
Study subjects and general methods	49
2.1 Study subjects	50
2.1.1 YDX study	50
2.1.2 Croatian participants in the Protein fucosylation in diagnosis of MODY study	52
2.1.3 Healthy subject group (Urinary steroids study)	53
2.2 DNA extraction and quantification	54
2.3 DNA amplification by polymerase chain reaction (PCR).....	55
2.4 DNA gel electrophoresis and imaging	57
2.5 DNA sequencing.....	58
2.5.1 Clean-up of PCR product	59
2.5.2 Sanger Sequencing	59
2.6 Genotyping.....	61
2.6.1 TaqMan PCR.....	62
2.6.2 TaqMan genotyping and analysis	63
2.7 Analysis of <i>HNF1A</i> variants.....	63
2.7.1 Mutation surveyor	63
2.7.2 Bioinformatics prediction tools.....	64
2.8 Functional studies of HNF1A protein variants	66
2.8.1 Site-directed mutagenesis (SDM)	66
2.8.2 Transformation of <i>Escherichia coli</i> and growth	69
2.8.3 Plasmid DNA extraction from transformed <i>Escherichia coli</i>	70
2.8.4 DNA quantification	71
2.8.5 Plasmid DNA sequencing.....	71
2.8.6 Maintenance of HeLa cell line and seeding technique	72
2.8.8 Transfection of HeLa cells with plasmid DNA.....	73

2.8.9 Transcription activity studies using dual luciferase reporter system	74
2.8.10 Harvesting and lysing HeLa cells for Western blotting and EMSA	75
2.8.11 Quantification of the total protein using Bradford assay	76
2.8.12 Protein expression using Western blot analysis	77
2.8.13 Assessment of DNA binding by Electrophoretic Mobility Shift Assay	79
2.9 Preparation of the plasma and serum samples and storage	54
2.10 Measuring C-peptide, GAD antibodies and CRP	81
2.10.1 C-peptide and high-sensitivity CRP assays	81
2.10.2 GAD antibodies	81
2.11 Release and labelling of N-glycans	82
2.11.1 Randomisation to correct for the batch effect	82
2.11.2 Sample preparation	83
2.11.3 2-aminobenzamide labelling	83
2.11.4 Clean-up of 2-AB labelled N-glycans and elution	83
2.12 HILIC-ultra performance liquid chromatography of N-glycans	84
2.13 Quality control of plasma pool samples & standard 2 (dextran)	85
2.14 Integration of the chromatography peaks	86
2.15 Gas chromatography mass spectrometry of urinary steroids	87
2.16 Statistical analysis	88
Chapter 3	89
<i>HNF1A</i> sequencing from the subjects with young adult onset non-autoimmune diabetes and interpretation of the <i>HNF1A</i> variants	89
3.1 Introduction	90
3.1.1 Assigning disease causality to the genetic variants	90
3.1.2 Known rare alleles causing <i>HNF1A</i> -MODY and common <i>HNF1A</i> alleles	95
3.1.3 <i>HNF1A</i> variants implicated in an increased risk of T2D	98
3.1.4 Aims of the study	100
3.2 Study subjects and methods	100

3.2.1 Study subjects	100
3.2.2 DNA extraction and quantification	102
3.2.3 DNA amplification and gel electrophoresis	104
3.2.4 Sanger sequencing and annotation of the variants	105
3.2.5 Validation of the sequencing results by genotyping.....	106
3.2.6 An assessment of the variants using bioinformatic prediction tools.....	108
3.2.7 Calculating the prevalence of HNF1A-MODY for the Croatian population	109
3.3 Results	109
3.3.1 <i>HNF1A</i> variants found in the study.....	109
3.3.2 Bioinformatic prediction for rare non-synonymous <i>HNF1A</i> variants	114
3.3.3 Review of previous reports in literature.....	115
3.3.4 Co-segregation of the rare non-synonymous <i>HNF1A</i> alleles and diabetes....	116
3.3.5 Clinical features of subjects with rare non-synonymous <i>HNF1A</i> variants.....	118
3.3.6 Allele frequency of the rare non-synonymous <i>HNF1A</i> alleles in large-scale sequencing samples.....	120
3.3.7 Prediction of the phenotypical effect of rare <i>HNF1A</i> alleles based on aspects presented	124
3.3.8 Prevalence of HNF1A-MODY in Croatia.....	127
3.4 Discussion.....	128
3.5 Conclusions	132
Chapter 4.....	133
Functional studies of uncharacterised HNF1A protein variants	133
4.1 Introduction	134
4.1.1 Experimental characterisation of rare non-synonymous <i>HNF1A</i> variants reported in literature	134
4.1.2 Aims of the experiments	139
4.2 Methods	139
4.2.1 Site-directed mutagenesis, transformation using <i>E.coli</i> and transfection of <i>HeLa</i> cells	141

4.2.2 Transcriptional activity using dual luciferase reporter system	143
4.2.3 HNF1A protein expression by chemiluminescent western blotting	145
4.2.4 DNA binding using Electrophoretic mobility shift assay	147
4.2.5 Statistical analysis	148
4.3 Characteristics of HNF1A protein variants undergoing functional assessment....	148
4.3.1 Features of <i>HNF1A</i> variants undergoing functional characterisation.....	149
4.3.2 HNF1A variants used as controls	155
4.4 Results	157
4.4.2 Transcriptional activity (TA) of the HNF1A protein variants using the dual luciferase reporter system	157
4.4.3 HNF1A protein expression	160
4.4.4 DNA binding of HNF1A protein variants using Electrophoretic mobility shift assay	162
4.4.5 Summary of the functional studies performed.....	165
4.4.6 Updated functional prediction score of <i>HNF1A</i> alleles	166
4.5 Discussion.....	168
4.6 Conclusions	173
Chapter 5.....	174
Biomarkers of HNF1A-MODY and translational follow-up	174
5.1 Introduction	175
5.1.1 Previously reported biomarkers of HNF1A-MODY	175
5.1.3 Aims of the study.....	177
5.2 Study subjects and methods.....	178
5.2.1 Study participants and inclusion criteria	178
5.2.2 Subject characteristics.....	178
5.2.3 <i>Sample size calculation for the assessment of biomarkers</i>	181
5.2.4 DNA extraction and sequencing	182
5.2.5 Assigning disease causality to <i>HNF1A</i> alleles	183

5.2.6 C-peptide, GAD antibodies and CRP assays.....	183
5.2.7 Analysis of plasma N-glycans	184
5.2.8 Statistical analysis	185
5.3 Results	187
5.3.1 High-sensitivity CRP in subjects with rare <i>HNF1A</i> variants.....	187
5.3.2 Plasma N-glycan profile in subjects with rare <i>HNF1A</i> variants	189
5.3.3 Discriminating HNF1A-MODY from young-adult onset type 2 diabetes using plasma glycans and hsCRP	Error! Bookmark not defined.
5.3.4 Clinical potential of GP30 and hsCRP.....	194
5.3.5 GP30 and CRP in predicting the functional effect of <i>HNF1A</i> variants	197
5.3.6 Clinical and biochemical predictors of HNF1A-MODY	202
5.4 Discussion.....	204
5.5 Conclusions	209
Chapter 6.....	210
Urinary steroid profile in HNF1A-MODY patients	210
6.1 Introduction	211
6.1.1 Steroid metabolism genes downregulated in the hepatocytes of <i>Hnf1a</i> knock- out mice.....	211
6.1.2 11 β -hydroxysteroid dehydrogenases and their involvement in glucose metabolism.....	212
6.1.3 3 α - and 3 β -hydroxysteroid dehydrogenases	215
6.1.2 Aims of the study	216
6.2 Study participants and methods.....	216
6.2.1 Study subjects	216
6.2.2 Study sample.....	217
6.2.3 Methods	218
6.3 Results	222
6.3.1 Subject characteristics.....	222

6.3.2 Urinary steroid metabolites in HNF1A-MODY.....	225
6.3.2 Enzymes with altered activity in HNF1A-MODY.....	226
6.4 Discussion.....	238
6.5 Conclusions	243
Overall discussion	244
References	257

Chapter 1

General introduction

The contents of this chapter have been published in the peer-reviewed journals:

- *Juszczak A, Owen K. Identifying subtypes of monogenic diabetes. Diabetes Management 2014; 4(1): 49-61.*
- *Juszczak A, Pryse R, Schuman A, Owen KR. When to consider a diagnosis of MODY at the presentation of diabetes: aetiology matters for correct management. Br J Gen Pract. 2016 Jun;66(647):e457-9.*

1.1 Young adult onset diabetes

Diabetes mellitus is a common chronic disease and a recognised public health concern worldwide. The global prevalence of diabetes in adults has risen from 4.7% in 1980 to 8.5% in 2014, which accounts for estimated 422 million people affected worldwide (WHO Global report on diabetes 2014). The increasing prevalence of diabetes is a challenge for health services, especially the more common type 2 diabetes (T2D), which is increasingly diagnosed at a younger age and implies rising costs due to a longer duration of treatment of the disease. According to International Diabetes Federation 2015 report, 12% of global health expenditure (around 673 billion dollars per annum) is spent on diabetes.

Type 1 diabetes (T1D) is the most common form of diabetes in children accounting for more than 95% of all diabetes in this group. The incidence of T1D is highest in the 10-15 year group (26.5 cases/100,000 of general population in the UK) and 5-8/100,000 people at the age of 15-19 years (National Paediatric Diabetes Audit Report 2013-14 and 2014-15, 2016, UK), thereafter falling sharply after the age of 15 years.

In the last two decades, the increasing prevalence of obesity has resulted in the increasing prevalence of T2D also in young adults, which had been reported as 2% of all children and young adults (<20 years old) in England and Wales, and 12.2% in the 20-24 years group (National Paediatric Diabetes Audit 2014-15). This means that the clinical distinction between T1D and T2D is ever more difficult to make and more patients will fall into the “atypical diabetes” category. However, a correct diagnosis is crucial as it guides

the treatment choice and impacts on future diabetes complication rate and therefore prognosis.

1.2 Monogenic diabetes

Along with an increasing prevalence of T2D in children and young adults, there are other causes of diabetes with onset in early adulthood that need to be considered. Monogenic diabetes is an inherited condition caused by single gene functional variants, and is comprised of an increasing number of subtypes due to a shift from a clinical to a molecular diagnostic classification. The biggest groups of monogenic diabetes are maturity onset diabetes of the young (MODY), neonatal diabetes and mitochondrial diabetes.

Maturity onset diabetes of the young was first described clinically by Tattersall as an insulin independent inherited form of diabetes (Tattersall 1974). The causal genes for MODY subtypes started emerging in the 1990s (Hattersley et al. 1992, Yamagata et al. 1996, Horikawa et al. 1997), and prompted the development of diagnostic molecular testing and therapeutic recommendations (Ellard et al. 2008).

More than 25 genes have been identified in patients with monogenic diabetes (including 12 associated with a MODY phenotype, Table 1.2.1), although functional alleles in most of them are very rare (<1% of MODY) and some reported in fewer than 3 families (Molven and Njolstad 2011).

Gene	Protein	Type of diabetes	Characteristic features	Reference
<i>HNF1A</i>	Hepatocyte nuclear factor 1 α	MODY	Progressive β -cell dysfunction, low renal threshold for glucose, sensitive to sulphonylureas (SU)	Yamagata et al. 1996
<i>HNF4A</i>	Hepatocyte nuclear factor 4 α	MODY	Progressive β -cell dysfunction, neonatal macrosomia/hypoglycaemia, sensitive to SU	Yamagata et al. 1996
<i>GCK</i>	Glucokinase	MODY	Stable mild hyperglycaemia, HbA1c <7.8%; treatment required only in pregnancy if baby doesn't have mutation, no complications	Hattersley et al. 1992
<i>HNF1B</i>	Hepatocyte nuclear factor 1 β	MODY	Progressive β -cell dysfunction, associated genito-urinary abnormalities, pancreatic exocrine dysfunction, abnormal LFTs, low magnesium	Horikawa et al. 1997
<i>ABCC8</i>	Sulphonylurea receptor 1	PNDM, TNDM, MODY	Common form of PNDM, decreased birth weight, responsive to high dose of SU, can also cause MODY phenotype	Babenko et al. 2006
<i>KCNJ11</i>	Kir6.2 unit of ATP-dependent K channel	PNDM, TNDM, MODY	Most common PNDM, decreased birth weight, associated developmental delay and epilepsy, responsive to SU, can also cause MODY	Gloyn et al. 2004
<i>INS</i>	Insulin	PNDM, MODY	10-15% of PNDM, 80% <i>de novo</i> variants, decreased birth weight, also reported to cause MODY	Edghill et al. 2008
<i>Mitochondrial DNA (most common m.3243A>G)</i>	Mitochondrial transcription terminating factor (MTTF)	MIDD	Progressive β -cell dysfunction, 75% sensorineural deafness, rarely other neurological features as MELAS, MERRF, Kearne-Sayre syndromes; myopathy, renal impairment, cardiomyopathy	Goto et al. 1990
<i>IPF1</i>	Insulin promoter factor 1	PNDM, possible MODY	Rare, pancreatic agenesis in homozygous variants, adult onset diabetes with increased weight and hyperinsulinism	Stoffers et al. 1997
<i>NEUROD1</i>	Neurogenic differentiation factor 1	PNDM, possible MODY	Rare, PNDM associated cerebellar hypoplasia, sensorineural deafness, developmental delay and visual impairment; adult onset DM with T2D phenotype	Malecki et al. 1999
<i>CEL</i>	Carboxyl ester lipase	Possible MODY	Reported in 2 families, childhood-onset pancreatic exocrine dysfunction/atrophy, adult onset diabetes (secondary?)	Reader et al. 2006
<i>WSF1</i>	Wolframin	PNDM, also reported as MODY	Associated with diabetes insipidus, optic atrophy and deafness	Bonnycastle et al. 2013
<i>KLF11</i>	Kruppel-like factor 11	Possible MODY	Reported in 3 families, poor evidence for causing MODY	Neve et al. 2005
<i>BLK</i>	B lymphoid tyrosine kinase	Possible MODY	Reported in 3 families, overweight subjects, poor evidence for causing MODY	Borowiec et al. 2009

Table 1.2.1 Genes reported as causing monogenic diabetes of young adult onset. MODY, maturity onset diabetes of the young; PNDM, permanent neonatal diabetes mellitus; TNDM, transient neonatal diabetes mellitus; MIDD, maternally inherited diabetes and deafness; hsCRP, high sensitivity C-reactive protein; SU, sulphonylurea; LFTs, liver function tests; AR, autosomal recessive; MELAS, mitochondrial diabetes, encephalopathy, lactic acidosis and stroke-like episodes; MERRF, myoclonic epilepsy, ragged red fibres in the muscles. Adapted from (Juszczak 2014).

Initially, MODY was estimated at around 0.1-1.8% in the whole diabetes population (Panzram and Adolph 1981, Ledermann 1995) and the prevalence of MODY in the United Kingdom was calculated as a minimum of 68-108 cases per million of general population (Shields et al. 2010) which corresponds to 0.1-0.15% of people with diabetes in the UK. The prevalence of HNF1A-MODY as the most common type of monogenic diabetes was reported as 0.07-0.4% of all diabetic population (Eide et al. 2008, Shields, Hicks et al. 2010, Kropff et al. 2011). A study of young adults diagnosed with diabetes below 45 years of age who were C-peptide positive and GAD antibody negative (n=297) reported the proportion of HNF1A-MODY as 4.6% in that group (Thanabalasingham et al. 2012). This proportion was overestimated as 2/12 *HNF1A* variants found in that study (p.G31D and p.G606S) are no longer regarded as MODY-causing and revised prevalence would be 3.4% in that population.

Data from the UK MODY diagnostic centre, published in 2010, reported 67% of MODY patients to have disease-causing alleles in transcription factor genes, hepatocyte nuclear factor 1A (*HNF1A*), 4A (*HNF4A*) and 1B (*HNF1B*). *GCK* alleles were responsible for 32% of MODY cases (Shields, Hicks et al. 2010). The reported prevalence of MODY subtypes varies between countries, reflecting the routine screening for diabetes performed in countries such as Germany, Italy, France and Spain, where GCK-MODY predominates (Estalella et al. 2007, Lorini et al. 2009). This is related to the fact that individuals with GCK-MODY are asymptomatic.

Establishing a genetic diagnosis is crucial as this is likely to be followed by change in diabetes treatment of individuals with HNF1A- and HNF4A-MODY, who are sensitive to sulphonylureas (SU). Individuals with MODY caused by alleles in *ABCC8* and *KCNJ11* are also responsive to SUs, which are recommended first line treatment in these subtypes of MODY. In GCK-MODY, treatment is generally not recommended as those individuals have only mild elevation of glucose and do not develop diabetes complications (Steele et al. 2014).

Individuals presenting with diabetes in adolescence or as young adults and who are atypical for type 1 or type 2 diabetes in phenotype should be considered for genetic testing starting with *HNF1A* and *HNF4A*. Characteristic features will guide towards *GCK*, *HNF1B* and mitochondrial diabetes which are described below. If these features are absent but a strong suspicion of MODY still remains, many genetic centres now offer a panel of the remaining monogenic diabetes genes. The next sections will describe the most common types of monogenic diabetes in more detail.

1.2.1 HNF1A-MODY

HNF1A-MODY is the most common type of monogenic diabetes in adults and estimated in 52% of individuals with MODY in the UK (Shields, Hicks et al. 2010). *HNF1A* encodes a transcription factor which is crucial for β -cell differentiation and function, and affected individuals have progressive β -cell defect. Variant alleles in *HNF1A* gene with established evidence that they can cause MODY phenotype are usually highly penetrant, as 96% of affected individuals develop diabetes by 55 years of age (Shepherd M 2001). On the other side of the spectrum are *HNF1A* alleles which have little or no functional impact (Flannick et al. 2013). More than 400 *HNF1A* variants had been reported in over 1200 families by 2013 (Colclough et al. 2013) and I discuss the *HNF1A* variants together with the characteristics of the HNF1A protein in section 1.4 and chapter 3 in more detail.

Individuals with disease-causing *HNF1A* alleles initially present with post-prandial glycosuria several years before diagnosis of diabetes. Those subjects have a lower renal threshold for glucose excretion (Bingham et al. 2001, Stride et al. 2005). This phenomenon results from failure to reabsorb glucose due to a decreased expression of the sodium-glucose co-transporter-2 (SGLT2) in the proximal renal tubule for which HNF1A is a transcription factor (Pontoglio et al. 2000).

HNF1A variant carriers initially have normal fasting glycaemia and subsequently develop an impaired glucose tolerance during OGTT with 2 hour glucose usually increased by more

than 4.5mmol/L from baseline (Stride et al. 2002). With time, fasting glucose gradually increases.

Individuals with HNF1A-MODY respond well to sulphonylurea (SU) treatment, and this clinical observation had been confirmed by Pearson *et al.* in a randomised crossover trial including 18 subjects with HNF1A-MODY and 18 with T2D matched for BMI and fasting plasma glucose (Pearson et al. 2003). They reported a 5-times greater fall in fasting glucose to gliclazide than metformin in subjects with HNF1A-MODY, and a 4-times greater response of glucose levels to gliclazide in individuals with HNF1A-MODY compared to those with type 2 diabetes. Therefore, sulphonylureas are recommended as first line treatment for subjects with HNF1A-MODY, who are typically well controlled on 20-40mg of gliclazide or 1.25-2.5mg glibenclamide daily. The alternative treatment is glinides, which also provide good control in this subtype of MODY (Tuomi et al. 2006).

Individuals initially misdiagnosed with type 1 diabetes can safely stop insulin and start low dose SU provided their C-peptide remains in the normal range. In the study by Shepherd *et al.* 79% (n=34) of subjects with an initial label of T1D and subsequently confirmed HNF1A-MODY successfully replaced insulin with gliclazide (Shepherd et al. 2009). Seventy one percent of them remained off insulin 39 months later with median HbA1c of 6.9%. Many individuals with HNF1A-MODY maintain good glycaemic control on this therapy for years, although at some stage secondary SU failure may occur and other hypoglycaemic agents will be required. There is no evidence base for a particular second-line therapy and any oral hypoglycaemic agent can be used if C-peptide remains in the normal range.

Rates of micro- and macrovascular complications were initially reported to be similar in the individuals with HNF1A-MODY and type 1 or type 2 diabetes matched for glycaemic control but not for duration of diabetes (mean 17 years in HNF1A-MODY group, 23 years in T1D and T2D) (Isomaa et al. 1998). The exception was IHD which prevalence was increased when compared to the individuals with T1D but decreased comparing to those with T2D. Steele *et al.* reported an increased all-cause mortality (hazard ratio 1.9) and cardiovascular mortality (HR adjusted for gender and smoking 2.3) in 153 subjects from 39 families with HNF1A-MODY when compared to unaffected family members (Steele et al. 2010). This study would be more relevant if individuals with HNF1A-MODY were compared to subjects with T1D or T2D matched for diabetes duration.

Oxford MODY clinic data on 49 subjects with HNF1A-MODY and 78 individuals with T2D matched for age and diabetes duration (17.1 vs 14.5 years, $p=0.26$) suggest that patients with HNF1A-MODY had a significantly lower rate of all vascular complications compared to subjects with T2D. A recently published study from Ireland on 60 subjects with HNF1A-MODY and 60 individuals with T1D matched for age, BMI and duration of diabetes also reported decreased rate of vascular complications in HNF1A-MODY subjects (Bacon et al. 2016). Those findings are likely due to better glycaemic control on SU treatment.

Subjects with HNF1A-MODY also have a better lipid profile compared to those with type 2 diabetes (McDonald et al. 2012). In the study by McDonald *et al.*, 267 patients with HNF1A-MODY had significantly higher HDL-cholesterol when compared with 297 individuals with T2D matched for gender and BMI (mean HDL 1.47 vs 1.15mmol/L,

p=0.0008). There was no significant difference in the TG levels between groups. When HNF1A-MODY subjects were compared to the matched healthy controls there was no difference in the levels of both lipid fractions. Nevertheless, if dyslipidaemia present, individuals with HNF1A-MODY should be considered for lipid lowering agents despite lack of evidence for disease specific cut-off.

1.2.2 HNF4A-MODY

Inactivating alleles in *HNF4A* also cause slowly progressing β -cell failure and share many common clinical features with functional allele which can cause HNF1A-MODY. It is the 3rd most frequently diagnosed type of MODY, estimated in 10% of MODY subjects in the UK (Shields, Hicks et al. 2010).

HNF4A plays an important role in the transcription factor network regulating pancreatic development and function (Odom et al. 2004). In the β -cell, HNF1A activates *HNF4A* via the P2 promoter, which then activates *HNF1A* transcription (Hansen et al. 2002, Boj et al. 2009). Therefore, a similar phenotype of patients with HNF1A- and HNF4A-MODY is not surprising. Individuals with HNF4A-MODY have an insulin sensitivity similar to non-diabetic subjects and similarly to the subjects with HNF1A-MODY are sensitive to sulphonylureas, which is the recommended first line treatment (Pearson et al. 2005). They are at risk of vascular complications, especially if their diabetes is poorly controlled.

Individuals with HNF4A-MODY had lower HDL-cholesterol when compared to unaffected family members (Pearson, Pruhova et al. 2005) which distinguishes them from subjects

with HNF1A-MODY. Another feature distinguishing HNF4A- from HNF1A-MODY is a normal renal threshold for glucose.

An increased rate of transient neonatal hypoglycaemia and macrosomia, even in the case of a non-diabetic mother, has been reported in subjects with *HNF4A-MODY* (Pearson et al. 2007). Pearson *et al.* reported on 54 babies with heterozygous *HNF4A* variants and compared them to 54 neonates from the family members which did not carry *HNF4A* variants. *HNF4A* variant carriers had a median birth weight 790g higher than unaffected family members and 56% of them had macrosomia (birth weight >4000g) comparing to 13% of unaffected family members. Eighty five *HNF1A* variant carriers were also examined and had a reported birth weight similar to the weight of unaffected family members.

Macrosomia in babies with *HNF4A* variants would be explained by foetal hyperinsulinism and insulin being a known growth factor. Neonatal hyperinsulinism was proven in three tested neonates with hypoglycaemia and *HNF4A* variants in that study. The hyperinsulinism phenotype of *Hnf4a* knock-out mice (Gupta et al. 2005) also supported this pathophysiology mechanism. Neonatal hypoglycaemia, defined as plasma glucose of <2.5mmol/L persistent beyond 24 hours after birth, was present in 14% (8/54) of neonates who carried *HNF4A* variants versus none in unaffected family members. Four of the subjects with neonatal hypoglycaemia inherited the *HNF4A* variant from the father which proves that hyperinsulinism was not induced by maternal hyperglycaemia. Only one baby with an *HNF1A* variant had documented persistent neonatal hypoglycaemia but his mother had diabetes during pregnancy.

The difference in the rate of macrosomia and neonatal hypoglycaemia between *HNF1A* and *HNF4A* carriers suggests that the function of HNF1A and HNF4A protein is independent in foetal life and changes later. The phenomenon of neonatal hyperinsulinism and diabetes later in life remains poorly understood.

1.2.3 HNF1B-MODY

HNF1B-MODY is caused by heterozygous inactivating variants or deletions in human transcription factor 2 gene (*TCF2/HNF1B*), which encodes hepatocyte nuclear factor 1 β . Although *HNF1B* is closely related to *HNF1A*, its disease-causing alleles result in a distinctive phenotype due to differences in timing and tissue-specificity of the expression. HNF1B has a role in the development of the pancreas, renal tract and Mullerian tract, explaining the range of associated abnormalities in affected individuals.

The co-existence of diabetes and extra-pancreatic features involving congenital genitourinary abnormalities should prompt testing for HNF1B-MODY. A systematic review of 50 studies including a total of 211 patients with *HNF1B* variants reported renal abnormalities in 89.6% of variant carriers and diabetes in 45% (Bellanne-Chantelot et al. 2005, Chen et al. 2010). The most commonly reported abnormalities involved renal cysts (renal cysts and diabetes syndrome (RCAD)), renal dysplasia, familial hypoplastic glomerulonephritic kidney disease, collecting system anomalies and less frequently female genital tract abnormalities such as bicornuate uterus. Most subjects with both the renal and diabetes phenotypes presented first with renal disease, and subsequently developed diabetes. The renal features also include hypomagnesaemia due to tubular magnesium wasting (Adalat

et al. 2009) and hyperuricaemia (Bingham et al. 2003). Pancreatic structural abnormalities (mainly atrophy) were reported in 10.4% of subjects with *HNF1B* variants and abnormal liver function in 15.2% of them (Bellanne-Chantelot et al. 2004, Chen, Gao et al. 2010).

When compared to BMI-matched individuals with HNF1A-MODY (n=30), subjects with HNF1B-MODY (n=8) had fasting hyperinsulinaemia with 2.4-fold lower insulin sensitivity, lower HDL-cholesterol and increased triglycerides, which were at similar levels to type 2 diabetes (Pearson et al. 2004). Their response to sulphonylureas was also similar to type 2 diabetes subjects (n=10). Insulin requirements in HNF1B-MODY tend to be low (<0.45U/kg), despite insulin resistance. Hyperinsulinaemia was present in subjects with normal renal function, so it was not a consequence of reduced renal excretion of insulin (Pearson, Badman et al. 2004).

Metformin or pioglitazone were suggested as first line oral agents in HNF1B-MODY (Pearson, Badman et al. 2004), if renal function allows it. If satisfactory control is not achieved, insulin should be considered.

1.2.4 GCK-MODY

Glucokinase catalyses glucose phosphorylation in the pancreatic β -cell. By 2009, more than 600 *GCK* variants had been reported in more than 1400 families (Osbak et al. 2009). Inactivating heterozygous variants in *GCK* reset the glucose threshold for insulin secretion, resulting in mild fasting hyperglycaemia of 5.5-8mmol/L from birth. Subjects with GCK-MODY have HbA1c usually below 8%, IFCC 64 mmol/mol (Martin et al. 2008, Steele et al.

2013) and their glucose increment on the oral glucose tolerance test (OGTT) does not exceed 4.5mmol/L (Stride, Vaxillaire et al. 2002). Over the life time, there is a small increase in hyperglycaemia which is parallel to the age-related increase in the non-diabetic population.

Subjects with GCK-MODY are asymptomatic and usually diagnosed either incidentally or as a part of routine screening, for example during pregnancy. Around 50% of children with incidentally found hyperglycaemia in Czech Republic were found to have GCK-MODY (Feigerlova et al. 2006).

After *GCK*-MODY is confirmed, it is recommended that individuals with this subtype of diabetes should stop all anti-glycaemic treatment as observational data suggest that medications do not improve hyperglycaemia or HbA1c (Stride et al. 2014) and vascular complications of diabetes are not reported (Velho et al. 1997, Steele, Wensley et al. 2013). One exception remains in the case of pregnancy where the mother has GCK-MODY and the baby does not. In this scenario, macrosomia can develop as in any diabetic pregnancy due to increased foetal insulin production. However, when the *GCK* functional allele has been inherited by the foetus, growth is normal as the mild maternal hyperglycaemia is at an appropriate level for foetal insulin secretion. The genetic testing of the foetal *GCK* via amniocentesis is generally not recommended. If acceleration in abdominal circumference on the foetal USS is detected, the mother would be treated with insulin which would be stopped after delivery (Velho et al. 2000, Spyer et al. 2001).

Another situation when treatment with metformin could be considered in subjects with GCK-MODY is when HbA1c rises from the normal baseline in the context of obesity, suggesting that the patient has developed T2D in addition to GCK-MODY.

1.2.5 Maternally inherited diabetes and deafness (MIDD)

Maternally inherited diabetes and deafness caused by mitochondrial DNA alleles affects approximately 1% of subjects with diabetes and most commonly results from A to G substitution at position 3243 (m.3243A>G) of the mitochondrial DNA (Goto et al. 1990). Co-segregation of early onset diabetes and sensorineural deafness with a maternal family history of either of them should trigger testing for mitochondrial diabetes.

Individuals with MIDD have progressive β -cell dysfunction with normal insulin sensitivity (Iwanishi et al. 1995). Most subjects are lean and present insidiously in the 3rd-5th decade of life. Sensorineural hearing loss occurs in more than 75% of affected individuals and often precedes the diagnosis of diabetes. Neurological involvement in carriers of disease-causing mitochondrial variants varies from mild neurological features to rare syndromes such as MELAS (mitochondrial diabetes, encephalopathy, lactic acidosis and stroke-like episodes), MERRF (myoclonic epilepsy, ragged red fibres in the muscles) or Kearne-Sayre syndrome (ophthalmoplegia, retinopathy, proximal myopathy, cerebellar ataxia) (Murphy et al. 2008). Individuals with MIDD have increased prevalence of renal impairment (28%) most commonly, secondary to focal segmental glomerular sclerosis, which often manifests with proteinuria (Guillausseau et al. 2001). Cardiomyopathy and generalised myopathy are also associated features.

The range of phenotype in MIDD depends on the levels of variant mitochondrial DNA in different tissues, termed heteroplasmy. Initial treatment of diabetes in MIDD depends on the presentation, but subjects with MIDD usually progress through diet and oral agents to insulin more rapidly than those with T2D. There is a consensus that metformin should not be used when individuals with MIDD have neurological features due to the increased risk of lactic acidosis.

1.2.6 Other rare forms of MODY

Table 1.2.1 listed other genes reported to cause MODY. Variants in genes known to cause neonatal (presenting in the first 6 months of life) diabetes (*ABCC8*, *KCNJ11* and *INS*) have been reported as causing the MODY phenotype in several case studies (Edghill et al. 2008, Bonnefond et al. 2012, Ovsyannikova et al. 2016). If a baby presents with hyperglycaemia in first 6 months of life, it is most likely monogenic diabetes (Iafusco et al. 2002). Neonatal diabetes is divided into permanent and transient diabetes mellitus (PNDM and TNDM). *ABCC8* encodes the SUR1 subunit and *KCNJ11* the Kir 6.2 subunit of the β -cell ATP-dependent potassium channel (Hartemann-Heurtier et al. 2009). Some family members of babies with neonatal diabetes secondary to *KCNJ11* or *ABCC8* variant alleles were found to have adult onset diabetes caused by the same gene alteration (Yorifuji et al. 2005). Based on experience from neonatal diabetes secondary to both variants, sulphonylureas should be the first line treatment.

IPF1 (insulin promoter factor 1/*PDX1*) (Stoffers et al. 1997), *NeuroD1* (neurogenic differentiation 1/*BETA2*) (Malecki et al. 1999) and *PAX4* (paired box 4) (Plengvidhya et al.

2007) genes represent three other transcription factors, all of which control the development of the endocrine pancreas. Variants in these genes have been reported to co-segregate with young adult onset diabetes in 1-3 families each but with phenotype different than classical MODY, as many subjects with *IPF1* and *NEUROD1* variants had been reported to have an increased weight and hyperinsulinism.

1.3 Differential diagnosis of young adult onset diabetes

Individuals with MODY characteristically present in the 2-4th decade of their life, maintain endogenous insulin secretion (C-peptide in the normal range) several years from diabetes onset and usually have at least two generations' family history of diabetes. Typically, they do not have features of insulin resistance such as *acanthosis nigricans* or central obesity and have negative β -cell antibodies.

Different subtypes of MODY have characteristic clinical features which can guide subsequent investigations as discussed above (Table 1.2.1). An online MODY probability calculator has been developed to guide referrals for genetic testing for MODY and gives a pre-test probability of an individual having MODY based on clinical features (age at diagnosis and current age, gender, current treatment, BMI, HbA1c and parental history of diabetes). However, it has been designed only for those diagnosed with diabetes before 35 years of age (Shields et al. 2012).

1.3.1 Current challenges in diagnosing MODY

In any genetic disease, assigning the molecular diagnosis is essential as it allows an optimal treatment, enables cascade screening of the family members and improves future outcomes. It is reported that around 80% of MODY subjects are misdiagnosed as the more common type 1 or type 2 diabetes (Shields, Hicks et al. 2010).

Studies involving MODY individuals report typically a 10-15 year delay from diabetes onset to genetic diagnosis (Shields, Hicks et al. 2010, Thanabalasingham, Pal et al. 2012). This means that some people with MODY would have missed an opportunity for appropriate treatment as β -cell dysfunction progresses with time, especially if good glycaemic control has not been achieved. Lack of MODY awareness among primary care physicians, high costs of *HNF1A* sequencing (£350/sample in the UK) and variable availability of genetic testing all impact on the low diagnostic rate of HNF1A-MODY.

Current guidelines for genetic testing for MODY include patients with diabetes diagnosed before 25 years of age, strong family history of diabetes of any type and insulin independence (Ellard, Bellanne-Chantelot et al. 2008). However, about 50% of individuals with MODY will be missed using those criteria (Shields, Hicks et al. 2010, Bellanne-Chantelot et al. 2011).

Thanabalasingham *et al.* investigated age-extended diagnostic criteria for MODY in subjects with an initial diagnosis of T1D (n=247) or T2D (n=277) before 45 years of age and recruited from Oxfordshire primary and secondary care. Twenty subjects with the

initial diagnosis of T1D had a maintained endogenous insulin production (C-peptide >0.2nmol/L after at least 3 years from diagnosis, normal range 0.2-1.2nmol/L) with negative β -cell antibodies and underwent sequencing of *HNF1A* and *HNF4A*. Two subjects (0.8%) had *HNF1A* variants found but currently one of them (p.G31D) is no longer considered as MODY-causing. Individuals with an initial diagnosis of T2D before 30 years of age or older but without features of metabolic syndrome (n=80) also underwent *HNF1A* and *HNF4A* sequencing and 3.6% (10/277) of the subjects initially diagnosed with type 2 diabetes had a likely disease-causing allele in *HNF1A* (Thanabalasingham, Pal et al. 2012). Based on the fact that 41.6% of individuals diagnosed with MODY in this study were above 30 years of age, the authors suggested investigating for MODY those diagnosed with diabetes before 45 years of age in the clinical settings.

1.3.2 Individuals initially diagnosed with Type 1 Diabetes

The phenotype of subjects with type 1 diabetes is often similar to that of individuals with MODY as both are likely to be lean. The typical person with type 1 diabetes presents with polyuria and polydipsia with subsequent weight loss. However, some individuals with late-onset T1D may have less marked symptoms. Diabetic ketoacidosis is extremely rare in MODY, although it can still occur (Egan et al. 2015).

It is important to check for β -cell autoimmunity, preferably anti-GAD (glutamic acid decarboxylase) or IA-2 (insulinoma associated protein 2) antibodies. They are strongly positive in the majority of subjects with autoimmune diabetes (Verge et al. 1994). Individuals with MODY generally have negative β -cell antibodies but they may be present

in <1% of subjects which is similar to non-diabetic population (McDonald et al. 2011). This differed in a paediatric group of subjects with MODY from Germany, where positive antibodies to pancreatic β -cells were found in 17% of cases (Schober et al. 2009). The authors of this study also reported that 34% of paediatric subjects with T2D were positive for 1 out of 4 β -cell antibody types, which could be a result of low cut-off levels, which were not provided. Another reason for overestimating the autoimmunity of the subjects in the MODY group could be overestimating the disease causality of some alleles, which could be in fact benign and present in subjects with true T1D.

Persistent C-peptide production (C-peptide >0.2 nmol/L, normal range 0.2-1.2nmol/L), especially outside the 'honeymoon period' in subjects previously diagnosed with T1D should prompt investigations for monogenic diabetes or T2D. Endogenous insulin secretion will also be supported by low insulin requirement (<0.5Unit/kg/day) and the absence of ketoacidosis during periods of low compliance with insulin injections.

Urinary C-peptide to creatinine ratio (UCPCR) in a spot post-prandial urine sample was proposed as a non-invasive alternative to serum C-peptide in subjects with normal renal function (Besser et al. 2011). UCPCR discriminated effectively between individuals with HNF1A-/HNF4A-MODY and those with T1D (UCPCR >0.2nmol/mol, ROC curve C-statistic 0.98). However, it did not discriminate MODY subjects from those with T2D (C-statistic 0.64). Another limitation of this study was the inclusion of subjects with at least 5 years from diagnosis of diabetes. This is too long to wait to assign the correct diagnosis.

An autosomal dominant inheritance pattern of diabetes is less likely to be seen in subjects with type 1 diabetes, as only 2-4% of them have an affected parent (Tillil and Kobberling 1987). The usefulness of family history (FH) of diabetes in selecting subjects for *HNF1A* genetic testing from the group initially labelled as T1D was questioned by Lambert et al., as they found that only 1 in 18 individuals with three generations' FH of diabetes and negative β -cell antibodies had a likely disease-causing *HNF1A* allele (Lambert et al. 2003). In clinical practice however, FH is used together with other clinical and biochemical features to select cases with high risk of MODY.

1.3.3 Individuals with initial diagnosis of type 2 diabetes

Subjects with type 2 diabetes are typically overweight with features of metabolic syndrome, and therefore insulin resistance. This includes central obesity, *acanthosis nigricans*, hypertension, non-alcoholic fatty liver, elevated triglycerides, low HDL-cholesterol and often polycystic ovary syndrome (Timar et al. 2000).

Most subjects with maturity onset diabetes of the young have a normal body mass index (BMI) and no features of metabolic syndrome. Individuals with inactivating *HNF1A* alleles were reported to have higher HDL-cholesterol than those with T2D (Owen et al. 2002, McDonald, McEneny et al. 2012). However, HDL-cholesterol of <1.12mmol/L had only 75% sensitivity and 64% specificity to differentiate both groups.

Sensitivity to small doses of sulphonylureas suggests HNF1A- or HNF4A-MODY and was discussed in the section 1.2.1.

Family history of diabetes is less helpful in distinguishing type 2 diabetes from MODY as the prevalence of parental diabetes is high in T2D, particularly with younger age at diagnosis. A case-control study of 44 subjects with HNF1A-MODY versus 44 age-matched individuals with type 2 diabetes did not report any significant difference in parental history of diabetes (65.9% vs. 63.6%; $p = 0.92$) (Owen, Shepherd et al. 2002). Similarly, in the SEARCH study of 586 diabetic subjects less than 20 years old, parental history of diabetes did not differ between participants with MODY and those selected for monogenic diabetes testing and found negative (50% vs. 51%) (Gilliam et al. 2013). Selection included subjects with C-peptide ≥ 0.8 ng/mL and negative β -cell antibodies. Thus the majority of patients found to be negative on testing for MODY had young onset T2D.

1.4 Hepatocyte nuclear factor 1 alpha (HNF1A)

I have chosen to focus my thesis on HNF1A-MODY as it is the most common type of monogenic diabetes in which assigning a correct genetic diagnosis provides an opportunity for treatment change resulting in better clinical outcomes. Thus, in this section, I will discuss features of the *HNF1A* gene and protein in more detail to provide background for *HNF1A* sequencing and functional studies of *HNF1A* variants described in chapter 3 and 4.

Hepatocyte nuclear factor 1 was discovered and named in 1987 when it was observed that it regulates a transcription of fibrinogen and alpha-1-antitrypsin in liver cells (Courtois et al. 1987). Bach *et al.* cloned the HNF1 protein from human cDNA library and assigned

it to chromosome 12q24.3 by *in situ* hybridisation (Bach et al. 1990); a linkage analysis in French families with the MODY phenotype mapped this MODY locus also to 12q (Vaxillaire et al. 1995). A year later, this locus was identified as the gene encoding a transcription factor *HNF1A* (Yamagata, Oda et al. 1996).

The *HNF1A* (previously *TCF1*, NC_000012.12) gene canonical transcript (ENST00000257555) consists of 10 exons, is 23,998bp in size (120978515..121002512, GRCh38.p7) and encoding a 631 amino-acid protein (Fig.1.3.1). The *HNF1A* promoter is 300bp long. There were 15 other transcripts described for *HNF1A* in Ensembl (Aken et al. 2016), of which only six translate to different length protein isoforms, depending on splicing. The main HNF1A protein isoforms are: A, containing 10 exons and predominantly expressed (mRNA) in foetal life, adult kidneys and liver; B, terminating in the distal part of exon 7 and C, truncating after exon 6 (Figure 1.4.1) (Harries et al. 2006). The latter two predominate in adult human islets.

HNF1A mRNA and protein expression was also reported in the gastrointestinal tract and mRNA low expression in the genital tracts and skin (Fagerberg et al. 2014). In addition, high expression of HNF1A protein was described in colorectal, liver, gastric and pancreatic cancers (www.proteinatlas.org) (Uhlen et al. 2015).

HNF1A protein (UniProtKB ID HNF1A_HUMAN, P20823) is a transcription factor for multiple genes regulating metabolism in the liver, kidneys and pancreatic islet cells during pancreatic development and in adult life (Ryffel 2001).

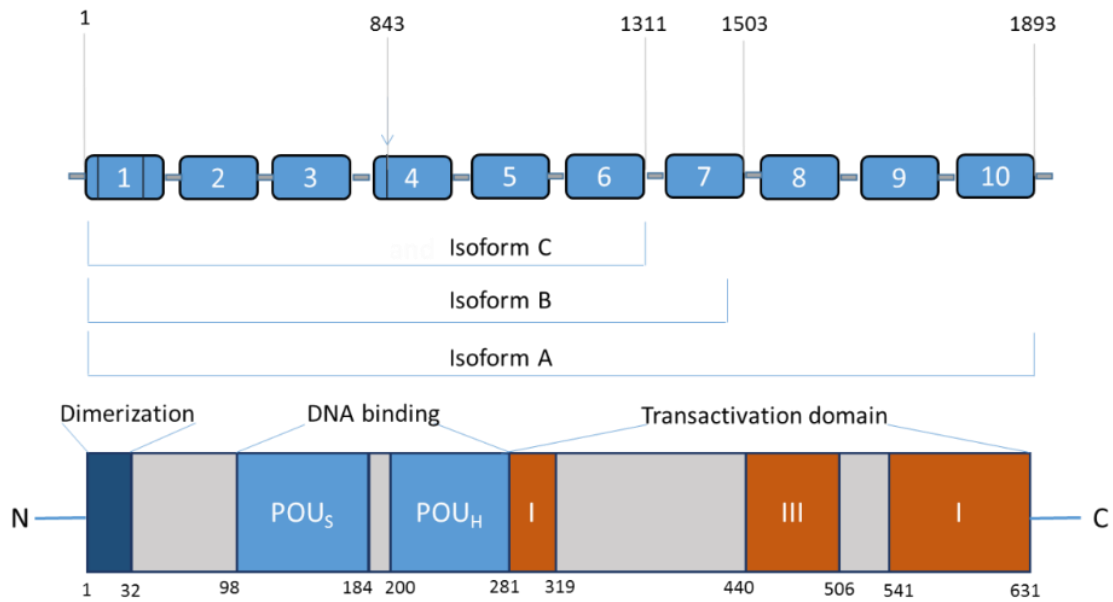


Figure 1.4.1 Structure of HNF1A gene (top) and corresponding protein (bottom) including 3 main isoforms. Numbers at the top correspond to nucleotide number in the gene and in inside blue boxes to the exon number; at the bottom to the amino acid number in the protein. POU_S, transcription factor specific homeodomain; POU_H, all transcription factors common homeodomain. Within transactivation domain: II and III, proline-rich regions; I, serine rich region.

Transcription factors are regulatory proteins, which control the rate of transcription of many genes by activating or suppressing them. HNF1A binds to at least 222 target genes in hepatocytes and 106 genes in pancreatic islets (Odom, Zizlsperger et al. 2004), including insulin (*INS*), pyruvate kinase (*PKLR*) and the glucose transporter 2 (*GLUT2/SLC2A2*) genes. As a transcription factor, hepatocyte nuclear factor 1-alpha binds to the inverted palindromic sequence 5-GTTAATNATTAAC-3.

HNF1A has three main functional regions: the N-terminal dimerization domain (1-32 amino-acids), DNA binding domain (98-281 AA) and transactivation domain (282-631 AA, Figure 1.4.1). The protein functions as a homodimer but it can also form heterodimers

with HNF1B, which shares 95% of its DNA binding domain and both can bind to the same DNA sequences.

A dimerization motif of HNF1A consists of 2 α -helices stabilised by a dimerization cofactor of HNF1-alpha, DCOH (Narayana et al. 2001), while the DNA binding domain is made of an universal homeodomain for several transcription factors (POU_H, made of 3 α -helices, 200-281 AA) and an HNF1A specific domain (POU_S, 5 α -helices, 98-184 AA, Figure 1.4.2) (Mendel and Crabtree 1991, Chi et al. 2002). The transactivation domain has 2 proline-rich (ADII and III, 29% of proline, Figure 1.4.1) and one serine-rich region (ADI, 26% of serine, 546-628 AA), which were reported as crucial for full transcription of the protein (Nicosia et al. 1990).

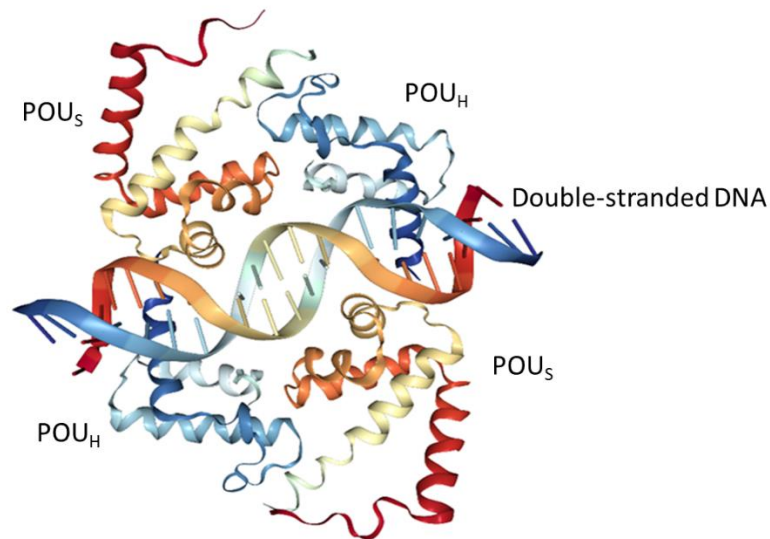


Figure 1.4.2 3D structure of DNA binding domain of HNF1A homodimer binding to DNA. Each DNA binding domain made of 8 α -helices: 3 α -helices forming a POU_H homeodomain (in shades of blue) and 5 α -helices making a POU_S homeodomain (in shades of yellow to red). Image downloaded from RCSB Protein Data Bank (Rose and Hildebrand 2015).

1.5 Challenges in interpretation of genetic variants

With an increase of large-scale sequencing projects and increased genetic testing in clinical settings, the interpretation of variant alleles remains challenging. This is especially related to missense variants, which could be neutral to the protein function and therefore not disease-causing, despite being non-synonymous.

The Exome Aggregation Consortium (ExAC) is a large-scale collaborative exome sequencing database of 60,706 unrelated individuals sequenced as part of different disease-specific and population genetic projects (Lek et al. 2016). Recently, it had been expanded to the Genome Aggregation Database (GnomAD) which contains data from 123,136 exomes and 15,496 whole-genomes (<http://gnomad.broadinstitute.org/>). Currently, GnomAD provides the largest dataset of human genetic variation and characterises minor allele frequency (MAF) depending on ancestry. There were 7.4 million alleles reported in ExAC individuals in total, which corresponds to an average of 1 variant allele per 8bp and 99% of the alleles had MAF <1%. Each exome in ExAC contained an average of 7.6 rare non-synonymous variants (MAF<0.1%) in well-characterised Mendelian disease genes, of which clearly not all cause penetrant autosomal dominant disease. The consequence of the greater sample size of ExAC was the reclassification of many variant alleles previously assigned disease-causing effect to benign (Walsh et al. 2017).

Finding rare missense coding variants in MODY genes, previously assumed to be causal for monogenic diabetes, in subjects with the type 2 diabetes phenotype or in middle-age healthy individuals started blurring genetic boundaries between MODY and T2D.

Flannick *et al.* aimed to characterise the spectrum of low-frequency variation (MAF<1%) in the main seven MODY genes in 4003 individuals from population-based cohorts (Flannick, Beer *et al.* 2013). They included subjects with and without diabetes (Framingham Heart Study (mean age in all groups 55.1-60.4 years, BMI 25.1-28.6kg/m²), Jackson Heart Study (mean age 55.5-60.5 years, BMI 31.3-33.3 kg/m²) and 771 individuals with extreme phenotypes (young lean subjects with T2D and obese non-diabetic individuals). The authors highlighted the limitations of previous databases of coding non-synonymous variants reported to cause MODY (such as Human Gene Mutation Database, HGMD) by showing that 1.5-1.8% of the individuals from population-based samples carry these variant alleles (including 12 variant alleles in *HNF1A*). Of those, 0.47-0.52% carried alleles which were present in HGMD classified as causal for MODY, evolutionarily conserved, predicted by bioinformatics as damaging and absent in 1000 Genomes project. Two thirds of HGMD-MODY variant carriers in this study did not have diabetes.

Most importantly, it has to be acknowledged the probable existence of a continuum of allelic functional effect, where the same variant alleles can result in different phenotypes from MODY to T2D, even in the same family, as other genetic, epigenetic and environmental factors also influence the penetrance of variant alleles (McCarthy 2017).

To avoid misinterpretation of genetic variants in clinical and research settings, guidelines for investigating causality of sequence variance in humans were published in 2014. The authors recommended providing all positive and negative evidence for variant implications; use of genetic, bioinformatics & experimental data if available; take advantage of public data sets of genomic variants such as ExAC; and not regarding previous reports as definitive (MacArthur et al. 2014). All listed aspects should be analysed together and neither of them is sufficient to assign disease causality of the variant allele in isolation. Previous evidence should be continuously re-evaluated when new knowledge is acquired (Quintans et al. 2014).

It was highlighted that when reviewing the frequency of the variant allele in large-scale sequencing studies an ancestry of an individual should be taken into account due to differences in MAF depending on ethnicity. Even if good evidence exists for the variant to cause the disease, full penetrance of the variant allele should not be assumed as other factors such as epigenetics, other variants/genes and environment can influence it.

It is recommended to validate the predicted effect of an allele experimentally using human cell lines or animal models. Yet, good experimental evidence that the particular variant affects protein function does not necessarily translate to clinical relevance due to the complexity of biological pathways in humans (Quintans, Ordonez-Ugalde et al. 2014).

Hence, while assessing *HNF1A* variants found in my study I considered all available evidence and performed functional studies for uncharacterised variants, which are discussed in chapters 3 and 4.

1.6 Biomarkers in HNF1A-MODY

A biomarker is an objectively measured and evaluated biological substance which is used to detect or monitor a physiological or pathological process or disease (Biomarkers Definitions Working 2001). An ideal biomarker should be easily quantifiable and inexpensive. In a rare genetic disease such as HNF1A-MODY, the aim is to find a biomarker that can select appropriate individuals for a definitive diagnostic test (DNA sequencing) and significantly reduce costs related to assigning a correct diagnosis. Multiple biological substances have been examined to establish their potential as biomarkers of HNF1A-MODY (Table 1.6.1) and most substances investigated were chosen due to the fact that HNF1A regulates their expression (Courtois, Morgan et al. 1987, Servitja et al. 2009), as discussed in section 1.4.4.

1.6.1 Previously investigated biomarkers of HNF1A-MODY

The mRNA of the gene encoding apolipoprotein M (ApoM) was downregulated more than 500-fold in liver cells from *Hnf1a* knock-out mice (Servitja, Pignatelli et al. 2009), so it seemed a good diagnostic biomarker candidate of HNF1A-MODY in humans. Mean ApoM levels were similar in subjects with HNF1A-MODY and T2D, while significantly lower in HNF1A-MODY when compared to T1D ($p=3.1 \times 10^{-18}$).

Biomarker	Sample	Study	Reason for investigation	Effect in HNF1A-MODY	Comment
1,5-anhydroglucitol	plasma	Skupien et al. 2008	Competing with glucose for a renal tubule transporter	Reduced compared to T2D	difference remained significant only if HbA1c 6.5-9%
Complement factor 5 and 8	serum	Karlsson et al. 2008	Regulated by HNF1A	Reduced compared to T2D	Both increased in uncontrolled hyperglycaemia
Transthyretin	serum	Karlsson et al. 2008	Regulated by HNF1A	Similar to T2D	Did not discriminate HNF1A-MODY patients
Apolipoprotein M	serum	Mughal et al. 2013	Regulated by HNF1A	Reduced compared to T1D, not T2D	Did not differentiate HNF1A-MODY from T2D
C-reactive protein (CRP)	serum	Thanabalasingham et al. 2011	Regulated by HNF1A	Reduced compared to T2D	Cheap and widely available but increased during inflammation
3-antennary fucosylated glycans	plasma	Thanabalasingham et al. 2013	Regulated by HNF1A	Reduced compared to T1D & T2D	Relatively expensive, no high throughput assay
Cystatin C	plasma	Nowak et al. 2013	Marker of GFR, correlated with CRP level	Similar to T1D & T2D	Did not discriminate HNF1A-MODY patients
MiRNA-103 and 224	serum	Bonner et al. 2013	Epigenetic regulation	Increased compared to healthy controls	Comparison with T1D & T2D required
MiRNA-103 and 224	urine	Bacon et al. 2015	Epigenetic regulation	MiRNA-224 increased compared to T2D, miRNA-103 similar	MiRNA-224 non-specific, increased in several types of cancer
Valine, glycine and betaine	urine	Gloyn et al. 2012	Metabolic profile of genetic disorders	Similar to other types of DM when corrected for glucose	Did not discriminate HNF1A-MODY patients
CD36	serum	Bacon et al. 2013	Elevated in subjects with T2D	Mildly decreased compared to T2D	Borderline significance (p=0.05)

Table 1.6.1 Previously investigated candidates for biomarker of HNF1A-MODY; GFR and their effect in subjects with HNF1A-MODY, glomerular filtration rate; MiRNA, microRNA.

Thus, it discriminated this type of monogenic diabetes from type 1, but not type 2 diabetes (Cervin et al. 2010, Mughal et al. 2013).

1,5-anhydroglucitol (1,5-AG) is a short-term marker of glycaemic control and was also investigated as a potential biomarker of HNF1A-MODY (Skupien et al. 2008). Mean 1,5-AG was lower in subjects with HNF1A-MODY (n=33) when compared to those with T2D (n=43, p=0.003) or non-diabetic individuals (n=47, p=5 × 10⁻⁶). However, the difference was significant only in subjects with HbA1c of 6.5-9% and 1,5-AG had a C-statistic of 0.88 for the cut-off <6.5µg/ml with sensitivity of 85.7% and specificity of 80% but only in subjects with the HbA1c in the mentioned range.

Karlsson *et al.* investigated complement 5, complement 8 and transthyretin as candidate biomarkers of HNF1A-MODY (Karlsson et al. 2008). The levels of complement 5 and 8 were similar in subjects with HNF1A-MODY (n=29) and non-diabetic controls (n=20) but around 40% lower than in the individuals with type 2 diabetes (n=20). Uncontrolled hyperglycaemia increased levels of complement components in general making them less useful in differentiating different types of diabetes. Levels of transthyretin, which is involved in thyroid hormones transport, were not significantly different in HNF1A-MODY when compared to T2D.

The potential of urinary metabolite profile to differentiate different types of diabetes using discovery platforms with UPLC-mass spectrometry and ¹H nuclear magnetic resonance spectroscopy was examined in subjects with HNF1A-MODY (n=25) versus

individuals with GCK-MODY (n=24), T2D (n=14) and healthy controls (n=34) (Gloyn et al. 2012). Despite the elevated valine, betaine, and glycine in the urine of HNF1A-MODY individuals when compared to the other types of diabetes using the discovery platforms, neither molecule had satisfactory discriminatory power when corrected for glucose level in the validation set (ROC derived C-statistics <0.60, HNF1A-MODY n=39, T2D n=158).

Cystatin C, which is a biomarker of glomerular filtration rate and its levels correlate with CRP levels, was found to be significantly lower in individuals with HNF1A-MODY (n=51) when compared to subjects with T1D and T2D (n=56 and 39 respectively, $p<0.05$). However, these findings did not replicate when validated in a bigger sample set and cystatin C was similar in patients with HNF1A-MODY (n=215) and other common types of diabetes (T1D n=58, T2D n=203, GCK-MODY n=170, HNF4A-MODY n=39 and HNF1B-MODY n=17) (Nowak et al. 2013).

Multiple research studies have investigated the diagnostic potential of short non-coding plasma microRNAs in different disorders. MiRNA-103 and 107 were reported to regulate insulin sensitivity (Trajkovski et al. 2011), while expression of miRNA-103 and 224 had been found to be 3.4- and 4.6-fold elevated in INS1 cells transfected with plasmid carrying the most common MODY-causing *HNF1A* allele, c.872dupC (Bonner et al. 2013). Circulating miRNA-103 and 224 were validated in 31 subjects with HNF1A-MODY and compared to 10 unaffected family members and 17 individuals with T2D (Bonner, Nyhan et al. 2013).

The median number of copies of miRNA-103 and 224 were 73- and 75-fold higher in subjects with HNF1A-MODY when compared to unaffected family members ($p=0.0001$ and 0.047 respectively) and miRNA-103 remained 8-times higher in comparison to the HbA1c-matched subjects with T2D ($p=3 \times 10^{-5}$).

The same research group assessed urinary levels of both microRNAs as a potential screening test for HNF1A-MODY and found that the levels of miRNA-224 were raised around 2-fold in subjects with HNF1A-MODY and type 1 diabetes (Bacon et al. 2015) when compared to healthy controls and T2D. The authors presented a C-statistic of 0.74 for discriminating HNF1A-MODY from healthy controls which is not useful in clinical practice where individuals with MODY need to be differentiated from other types of diabetes. MicroRNA-103 was elevated in HNF1A-MODY, type 1 and 2 diabetes. Thus neither miRNA discriminated HNF1A-MODY from both common types of diabetes.

Circulating CD36 was found to be elevated in subjects with insulin resistance and T2D. Therefore, it was hypothesised that it could be reduced in HNF1A-MODY subjects who are lean with good insulin sensitivity (Bacon et al. 2013). The levels of circulating CD36 were only slightly lower in HNF1A-MODY than in weight-matched individuals with T2D with borderline significance (mean 0.7 ± 0.4 vs 1.0 ± 0.5 , $p=0.05$).

1.6.2 Biomarkers of HNF1A-MODY with a translational potential

1.6.2.1 High-sensitivity CRP

C-reactive protein (CRP), similar to many other biomarker candidates, is regulated by HNF1A (Toniatti et al. 1990). Genome-wide association studies (GWAS), the Pharmacogenomics and Risk of Cardiovascular Disease and the Cardiovascular Health Studies reported an association between two common single-nucleotide polymorphisms (rs7953249 and rs735396) near *HNF1A* and plasma CRP levels (Reiner et al. 2008, Ridker et al. 2008, Okada et al. 2011). Owen *et al.* hypothesised that inactivating variants in *HNF1A* would lead to a decrease in CRP that could be of diagnostic utility (Owen et al. 2010).

The pilot study examined CRP measured by high-sensitivity assay (hsCRP) in 31 subjects with HNF1A-MODY, 24 with GCK-MODY, 316 with T1D, 240 with T2D and 198 non-diabetic controls from the UK (Owen, Thanabalasingham et al. 2010). Mean hsCRP was 0.2 (0.03-1.14) mg/L in HNF1A-MODY versus 1.33 (0.28-6.14) mg/L in T2D, 1.01 (0.19-5.33) mg/L in GCK-MODY and 0.58 (0.10-2.75) mg/L in T1D group (all $p \leq 0.009$). The receiver operating characteristic (ROC) curve derived C-statistic for differentiating HNF1A-MODY from T2D was 0.80 for CRP cut-off <0.4 mg/L and less discriminatory versus all other types of diabetes together (C-statistic 0.75).

The initial results were validated in a multi-centre study including patients from 7 European countries with HNF1A-MODY (n=457), GCK-MODY (n=404), HNF4A-MODY

(n=54) and young-onset T2D (n=582). The study replicated the initial results with hsCRP significantly lower in subjects with HNF1A-MODY when compared to all other types of diabetes (median CRP in HNF1A-MODY in all countries 0.03-0.25 (0.03-0.98) mg/L vs 1.44-5.21 (0.40-16.3) mg/L in T2D, $p < 1.3 \times 10^{-5}$) and distinguished better than the pilot study between HNF1A-MODY and young-onset type 2 diabetes with ROC derived C-statistics of 0.79-0.97 from different countries (Thanabalasingham et al. 2011). HsCRP was concluded to be a good biomarker for HNF1A-MODY and the cut-off $< 0.5 \text{ mg/L}$ had sensitivity and specificity of around 80% to distinguish this type of monogenic diabetes from young-onset T2D (Thanabalasingham, Shah et al. 2011). A disadvantage of using CRP is the increase of CRP in the acute phase reaction during a co-existing infection or inflammation. Therefore, it is recommended to repeat the test after several weeks if CRP is initially $> 10 \text{ mg/L}$.

The authors also reported that subjects with missense *HNF1A* variants had lower hsCRP than those with protein truncating variants (0.03 vs 0.08 mg/L, $p = 1.5 \times 10^{-5}$), and that individuals with variants in the dimerization and DNA binding domains had lower hsCRP than those with *HNF1A* variants in the transactivation domain (0.03 vs 0.1 mg/L, $p = 4.4 \times 10^{-6}$).

1.6.2.2 N-glycan profile in HNF1A-MODY

The N-glycosylation of proteins is essential for their post-translational modification in the endoplasmic reticulum and produces branches of oligosaccharides added to the nitrogen of the amine group of asparagine (N-glycans), which influence protein structure and function (Gornik et al. 2012). Severe glycan deficiencies affecting genes encoding the

enzymes involved in forming core glycan structures lead to autosomal recessive congenital disorders of glycosylation with multisystem defects (Freeze 2006). Several fucosyltransferases are involved in the fucosylation (adding fucose residue) of plasma proteins (Javaud et al. 2003).

Similar to the observation in CRP, two variants near *HNF1A* (rs7953249 and rs735396) were associated with variation in N-glycan levels which led to efforts to determine whether they could provide a potential biomarker of HNF1A-MODY (Lauc et al. 2010). This was further supported by the fact that HNF1A and HNF4A regulate GDP-fucose synthesis, upregulate antennary fucosyltransferases 3, 5 and 6, and downregulate core fucosyltransferase 8 in the liver and pancreas (Lauc, Essafi et al. 2010).

Thanabalasingham *et al.* hypothesised that subjects with heterozygous inactivating variants in *HNF1A* have reduced levels of antennary fucosylated N-glycan structures (Thanabalasingham et al. 2013). A multicentre study including participants with HNF1A-MODY (n=188), HNF4A-MODY (n=40), GCK-MODY (n=118), type 1 (n=98) and 2 diabetes (n=167) showed that the index derived from 3-antennary fucosylated plasma N-glycans ($DG9/(DG9+DG8)$, where DG states for desialylated glycan) was significantly decreased in HNF1A-MODY compared to T2D (0.09 (0.06-0.13) vs 0.25 (0.18-0.33) and type 1 diabetes (0.09 (0.06-0.13) vs 0.28 (0.20-0.34), $p \leq 1 \times 10^{-34}$ for both comparisons). The above index discriminated individuals with HNF1A-MODY from those with type 1 and 2 diabetes well with C-statistics of 0.91-0.94 (Thanabalasingham, Huffman et al. 2013). Similar to hsCRP, the DG9 index was lower in subjects with missense than with protein truncating *HNF1A*

variants, but the significance of this comparison was lost when only probands were included. Furthermore, the DG9 index was lower in individuals with missense variants in dimerization and DNA binding domains than in the transactivation domain. The difference remained significant after excluding relatives.

Table 1.6.2 compares CRP with DG9 glycan index as potential biomarkers of HNF1A-MODY.

	High-sensitivity CRP	Fucosylated N-glycan index
Price	low cost	no current price as research tool only, but likely to be expensive
Availability	widely available	not widely available
ROC C-statistics vs T2D	0.79-0.94	0.91
Sensitivity	83% for CRP <0.5mg/L	88% for DG9 index <0.16
Specificity	86% for CRP <0.5mg/L	81% for DG9 index <0.16
ROC C-statistics vs T1D	0.66	0.94
Sensitivity	not calculated as C-stat low	88% for DG9 index <0.16
Specificity	not calculated as C-stat low	88% for DG9 index <0.16
Affected by inflammation	elevated due to infection/inflammation	not affected

Table 1.6.2 Comparison of hsCRP and fucosylated N-glycan index as potential biomarkers; ROC, receiver operating characteristics; C-stat, C-statistics; DG9, desialylated glycan 9.

The above studies on CRP and N-glycan profile were designed to compare individuals with extreme phenotypes selected from research centres with an expertise in monogenic diabetes. This approach is likely to overestimate the diagnostic performance of the biomarker due to spectrum bias (Whiting et al. 2004). Therefore, to clarify the true performance of hsCRP and glycan profile as biomarkers of HNF1A-MODY, I evaluated

them in an unselected sample set resembling patients from a young adult diabetes clinic, and chapter 5 describes the results.

1.7 Adrenal steroids and *Hnf1a* knock-out mouse

1.7.1 Expression of steroidogenesis enzymes in *Hnf1a* knock-out mice

Humans with MODY due to heterozygous inactivating *HNF1A* variants have diabetes, while heterozygous *Hnf1a* (+/-) mice do not (Pontoglio et al. 1998). Complete *Hnf1a* knock-out (-/-) mice have reduced glucose-stimulated insulin release and hyperglycaemia (Pontoglio et al. 1996). Despite differences between rodent model and human, animal models provide useful insights into disease mechanisms and give background to the research in humans.

Several genes involved in adrenal steroid metabolism were found to be downregulated in the liver cells of the *Hnf1a* -/- mice (Servitja, Pignatelli et al. 2009) (Table 1.7.1). Of these, genes encoding three isoforms of 3 β -hydroxysteroid dehydrogenase (*Hsd3b2*, 3 and 5) catalyse one of the initial steps of steroidogenesis for gluco-, mineralocorticoids and adrenal steroids. 11 β -hydroxysteroid dehydrogenase type 1 (*Hsd11b1*) converts inactive cortisone to active cortisol and 17 β -hydroxysteroid dehydrogenase type 2 (*Hsd17b2*) facilitates conversion of androstenedione to testosterone. Aldo-keto-reductases (*Akr1c6* and *c20*) reduce the A-ring of multiple steroids.

Therefore, I hypothesised that some of the enzymes encoded by the above genes could have reduced activity in HNF1A-MODY subjects and that this might provide a potential

biomarker for HNF1A-MODY. An assessment of the activity of the above enzymes involved in steroid metabolism using urinary steroid profile is discussed in detail in chapter 6.

Gene	Protein	Fold change
<i>Hsd3b5</i>	hydroxysteroid dehydrogenase-5, delta<5>-3-beta	581.1
<i>Hsd3b2</i>	hydroxysteroid dehydrogenase-2, delta<5>-3-beta	35.3
<i>Akr1c6</i>	aldo-keto reductase family 1, member C6	17.5
<i>Cyp7b1</i>	cytochrome P450, family 7, subfamily b, polypeptide 1	17.2
<i>Hsd11b1</i>	hydroxysteroid 11-beta dehydrogenase 1	15.1
<i>Akr1c20</i>	aldo-keto reductase family 1, member C20	11.4
<i>Hsd17b2</i>	hydroxysteroid (17-beta) dehydrogenase 2	4.3
<i>Hsd3b3</i>	hydroxysteroid dehydrogenase-3, delta<5>-3-beta	3.9

Table 1.7.1 Genes involved in steroid metabolism and downregulated more than 2-fold in the liver cells of *Hnf1a* ^{-/-} mouse (Servitja, Pignatelli et al. 2009).

1.7.2 The role of glucocorticoids in glucose metabolism

Glucocorticoids are closely linked to glucose metabolism and their excess can cause metabolic syndrome and diabetes. Thus, changes in their metabolism in subjects with any type of diabetes may explain some aspects of the patients' phenotype.

In the physiological state, glucocorticoids (GC) and insulin have an opposite effect on hepatic gluconeogenesis. Therefore, GC stimulate and insulin inhibits phosphoenolpyruvate-carboxykinase (PEPCK) and glucose-6-phosphatase (Jitrapakdee 2012).

In the fasting state, GC increase levels of fatty acids by lipolysis in the adipose tissue (Gathercole et al. 2011). In the fed state, glucocorticoids promote carbohydrate and lipid

storage, differentiation of omental preadipocytes and in excess lead to a preferential increase of intra-abdominal adipose tissue and central obesity.

Increased levels of glucocorticoids cause insulin resistance and reduced glucose uptake in the target tissues with impaired glucose tolerance. A potent glucocorticoid, dexamethasone, reduced insulin binding to the insulin receptor in rat hepatocytes, which was dose- and time-dependent (Olefsky et al. 1975). Insulin stimulates tyrosine kinase activity of its receptor resulting in phosphorylation of its substrate, insulin receptor substrate-1 (IRS-1), which activates phosphatidylinositol 3-kinase (PI3K). This way insulin increases glucose uptake in muscles, adipose tissue and the liver. Dexamethasone-treated rats showed decreased phosphorylation of the insulin receptors in hepatocytes and decreased activity of PI3K in hepatocytes and muscles (Saad et al. 1993).

Cortisol is the main active glucocorticoid in humans and its availability in the target tissues is regulated by 11 β -hydroxysteroid dehydrogenase type 1 (11BHSD1), which was downregulated in *Hnf1a* $-/-$ mice. 11BHSD1 converts mainly inactive cortisone to cortisol (Tomlinson et al. 2004) but could be bidirectional. The opposite conversion of cortisol to cortisone is also catalysed by 11 β -hydroxysteroid dehydrogenase type 2. Blocking 11 β -hydroxysteroid dehydrogenase type 1 in rodents improved their glucose tolerance and insulin sensitivity (Alberts et al. 2002, Alberts et al. 2003).

Individuals with Cushing's syndrome, i.e. glucocorticoid excess, develop metabolic syndrome with central obesity, hypertension and insulin resistance with overt diabetes in

30-40% or impaired glucose tolerance in a further 20-30% of subjects (Friedman et al. 1996, Biering et al. 2000).

Individuals with HNF1A-MODY are usually slim and are not insulin resistant, thus altered steroid metabolism with a reduced glucocorticoid production could be anticipated but it has not been investigated before.

1.7.3 Steroidogenesis pathway

As I have analysed urinary steroid profile in subjects with HNF1A-MODY, T2D and healthy controls and chapter 6 discusses the results, it is essential to describe the main aspects of steroidogenesis and steroid metabolites.

Steroidogenesis occurs in humans mainly in the adrenal glands, testicular Leydig cells, ovarian granulosa cells and trophoblast cells in the placenta. All steroid hormones are derived from cholesterol, thus based on the 4-ring structure of cyclopentanophenanthrene. ACTH (adrenocorticotrophic hormone) stimulates cholesterol synthesis and the synthesis of all adrenal steroid hormones, while production of androgens is also stimulated by luteinising hormone (LH).

The enzymes involved in steroidogenesis are either in the cytochrome P450 (Pigment 450, absorbs light at 450nm, *CYP*) group of unidirectional enzymes mediating hydroxylation and carbon-carbon bond cleavage, or hydroxysteroid dehydrogenases (HSDs) which reduce ketosteroid or oxidase hydroxysteroid and could be bidirectional (Miller and Auchus 2011).

Some key steroidogenic enzymes catalyse several reactions in the pathway (Miller 1988).

An example is 17 α -hydroxylase type 1 (encoded by *CYP17A1*) (Figure 1.7.1).

The main mineralocorticoid in human is aldosterone, which as part of renin-angiotensin II-aldosterone axis. It controls fluid and electrolyte balance and thus blood pressure. 3 β -hydroxysteroid dehydrogenase type 2 (HSD3B2, downregulated in *Hnf1a* $-/-$ mice), 21 α -hydroxylase type 2 (CYP21A2) and 11 β -hydroxylase type 2 (CYP11B2) are involved in the aldosterone synthesis pathway (Figure 1.7.1) (Kawamoto et al. 1992).

The enzymes involved in the glucocorticoid pathway are HSD3B2, 21 α -hydroxylase type 2 and 11 β -hydroxylase type 1 (CYP11B1) (Fardella and Miller 1996).

Adrenal androgens can originate from different steroid precursors and this pathway is catalysed by HSD3B2 and 17 β -HSDs (17-ketosteroid reductases, HSD17B1-3) to finally produce testosterone. *Hsd17b2* was also downregulated in *Hnf1a* $-/-$ mice. In Leydig cells and the prostate, testosterone is metabolised to more potent dihydrotestosterone (DHT) by 5 α -steroid reductase type 2 (SRD5A2) and in the ovaries, placenta or adipose tissue to oestradiol by P450 aromatase (CYP19A1).

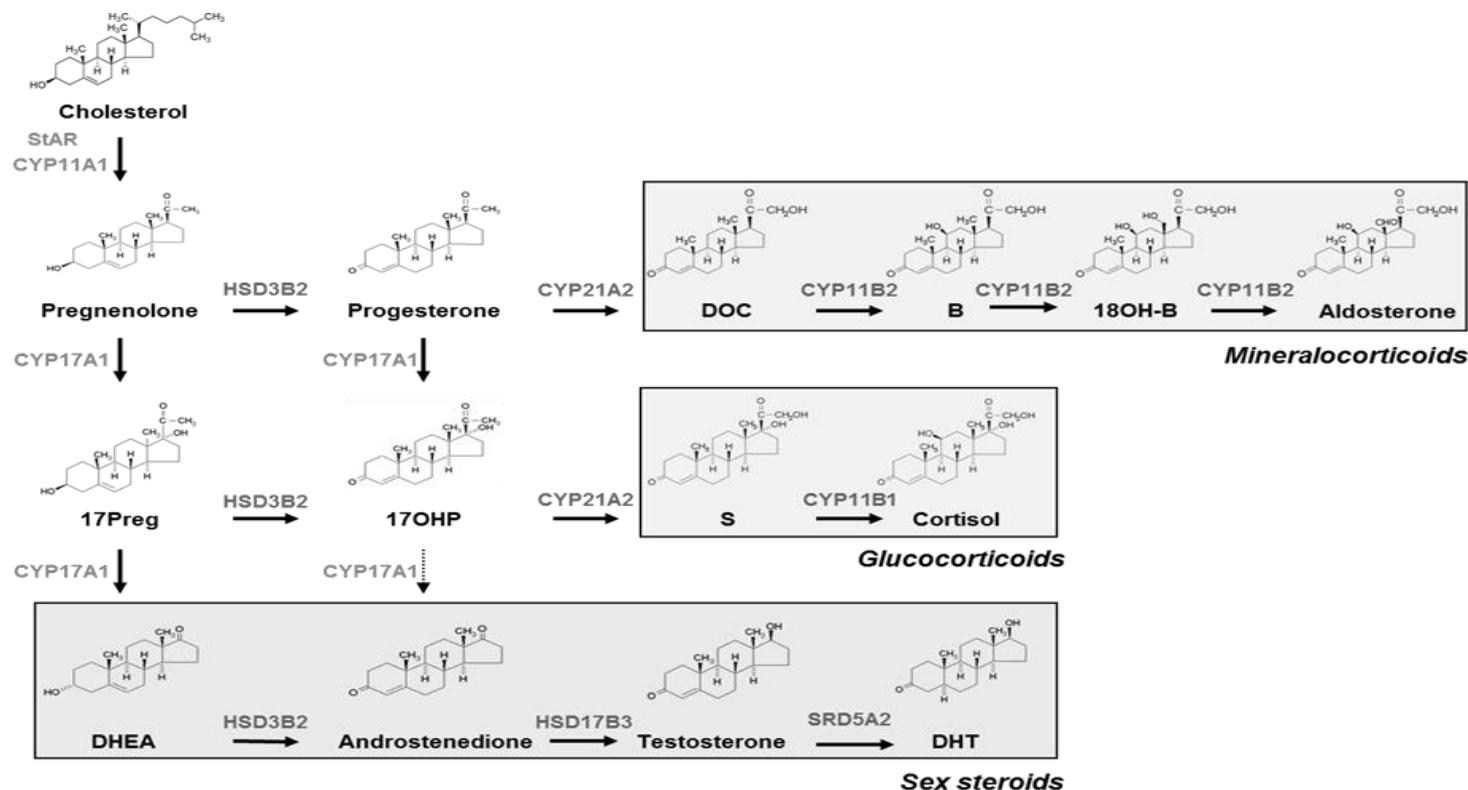


Figure 1.7.1 Adrenal steroidogenesis pathway; DHEA, dehydroepiandrosterone; HSD, hydroxysteroid dehydrogenase; OHP, hydroxyprogesterone; StAR, steroidogenic acute regulatory protein; DOC, 11-desoxycorticosterone; B, corticosterone; S, 11-desoxycortisol; DHT, dihydrotestosterone; CYP11A1, cytochrome P450 side-chain cleavage enzyme; CYP17A1, steroid 17-hydroxylase or 17/20-lyase; CYP11B1, steroid 11B-hydroxylase; CYP11B2, steroid 11/18B-hydroxylase; CYP21A2, 21-hydroxylase; HSD3B2, 3B-hydroxysteroid dehydrogenase type 2; HSD17B3 17B-hydroxysteroid dehydrogenase type 3; SRD5A2, 5A-steroid reductase type 2. Active steroids shaded; enzymes catalysing conversions given above arrows in grey.

1.7.4 Urinary steroid metabolites in human

Table 1.7.2 lists all adrenal steroid metabolites and Figure 1.7.2 places most of them in the context of steroidogenesis pathway. In clinical practice, urinary steroid profile has proved to be useful in the diagnosis of congenital adrenal hyperplasia (CAH), caused by deficiencies in different steroidogenesis enzymes (Caulfield et al. 2002), adrenocortical carcinoma (ACC) (Arlt et al. 2011) and in detecting the abuse of anabolic steroids (Gosetti et al. 2013).

Several enzymes in adrenal steroidogenesis catalyse multiple steps of the pathway; thus ratios have been derived to account for many substrates and products of each reaction (Wooding and Auchus 2013). The particular steroid metabolites and their ratios investigated in subjects with HNF1A-MODY are described in chapter 6.

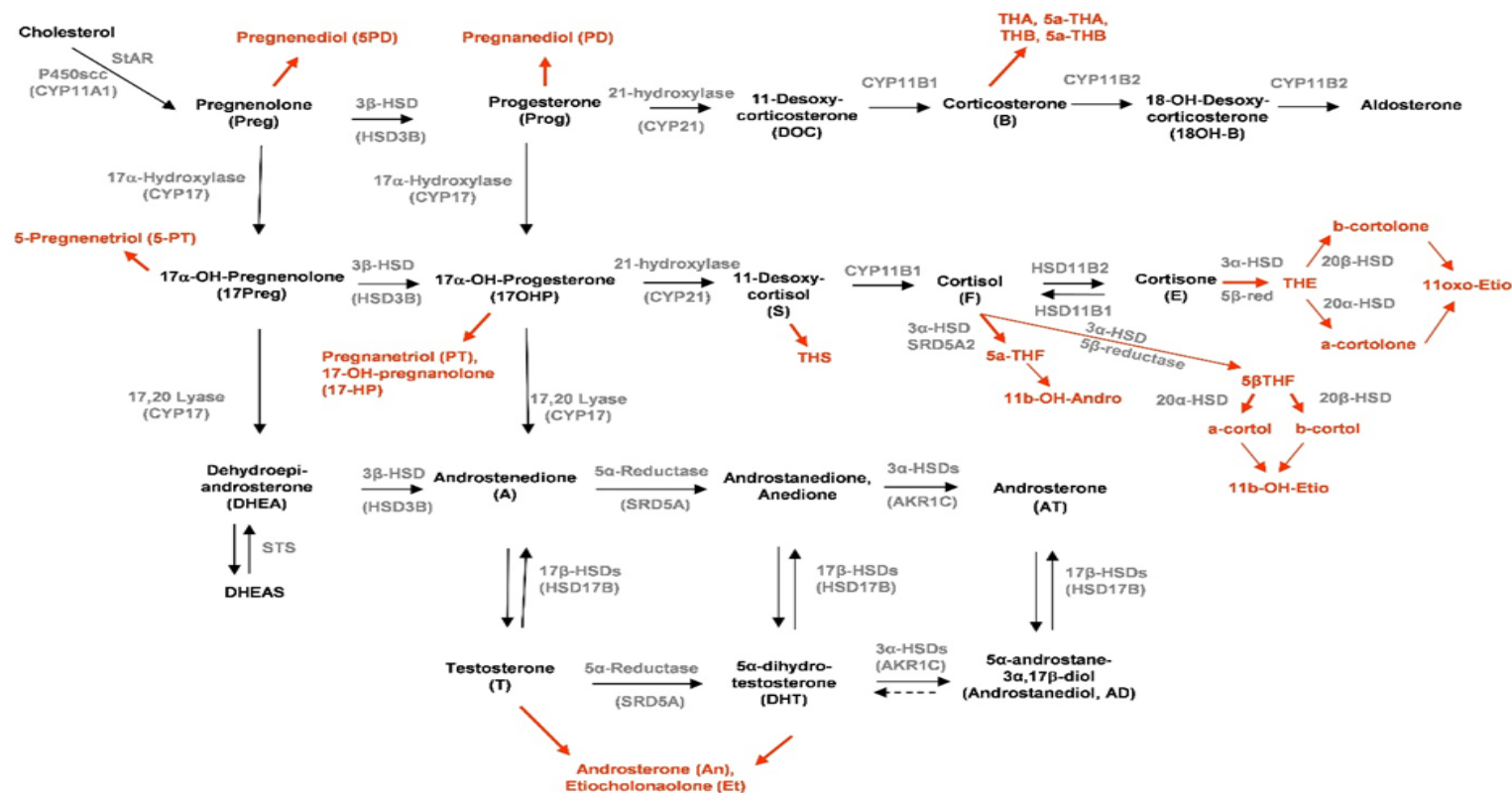


Figure 1.7.2 Steroidogenesis pathway including steroid metabolites. Adrenal steroids in black, adrenal steroid metabolites in red, enzymes catalysing steroid conversions in gray. CYP11A1, cytochrome P450 side-chain cleavage enzyme; CYP17A1, steroid 17-hydroxylase or 17/20-lyase; CYP11B1, steroid 11B-hydroxylase; CYP11B2, steroid 11/18B-hydroxylase; CYP21A2, steroid 21-hydroxylase; HSD3B2, 3B-hydroxysteroid dehydrogenase type 2; HSD17B3, 17B-hydroxysteroid dehydrogenase type 3; SRD5A2, 5A-steroid reductase type 2; AKR1C, 3A-hydroxysteroid dehydrogenase. Abbreviations of steroid metabolites expanded in Table 1.7.2.

Steroid group	Steroid precursor	Metabolites
Mineralocorticoids		
Mineralocorticoid precursors		THA (tetrahydro-11-dehydrocorticosterone)
		5 α -THA
		THB (tetrahydrocorticosterone)
		5 α -THB
		18-OH-THA
		TH-DOC (tetrahydro-11-desoxycorticosterone)
		3 α 5 β -TH-Aldosterone
Glucocorticoids		
Glucocorticoid precursors		PD (pregnanediol)
		17-HP (17-OH-prognanolone)
		3 α 5 β -17HP
		PT (pregnanetriol)
		PT'ONE (pregnanetriolol)
		THS (tetrahydro-11-desoxycortisol)
Cortisol		F (cortisol)
		6 β -OHF
		THF (tetrahydrocortisol)
		5 α -THF
		α -cortol
		β -cortol
		11 β -OH-Etiocholanolone
Cortisone		E (cortisone)
		THE (tetrahydrocortisone)
		α -cortolone
		β -cortolone
		11-oxo-Etiocholanolone
Androgens		
Androgen precursors		DHEA (dihydroepiandrosterone)
		16 α -OH-DHEA
		5-PT (5-pregnanetriol)
		5-PD (pregnenediol)
		Pregnandionol
		Total PD (5-PD+Pregnandionol)
Androgens		Andro (Androsterone)
		Etio (Etiocholanolone)
		11 β -OH-Andros

Table 1.7.2 Adrenal steroid hormone metabolites. All abbreviations expanded in the table and Figure 1.7.2.

1.8 Aims of the thesis

The introduction outlined the challenges in the differential diagnosis of young-adult onset diabetes and identified the need for a good screening tool to best select individuals for genetic testing for MODY. Increasing large-scale sequencing projects and clinical use of their data demand an improvement of the interpretation of reported genetic variation and functional assessment of uncharacterised variants to better inform assigning the functional effect of variant alleles.

Therefore, the objectives of my thesis are to:

1. Systematically assess all rare *HNF1A* variant alleles from an unselected population with young adult-onset non-autoimmune diabetes from the UK and Croatia (Chapter 3);
2. Perform functional studies for uncharacterised *HNF1A* variants and combine genetic and functional data to understand the clinical impact of selected variant alleles (Chapter 4);
3. Evaluate the performance of hsCRP and glycan profile as markers for HNF1A-MODY in an unselected population with young adult-onset non-autoimmune diabetes and assess their translational potential (Chapter 5);
4. Compare the urinary steroid profile in subjects with HNF1A-MODY, T2D and healthy controls, as well as establish if the urinary steroid profile in HNF1A-MODY

individuals can explain some aspects of their phenotype or identify a new biomarker of HNF1A-MODY (Chapter 6).

Chapter 2

Study subjects and general methods

2.1 Study subjects

2.1.1 YDX study

The Young Diabetes in Oxford (YDX) study was set up to enhance the clinical, physiological, epidemiological and molecular characterisation of subtypes of diabetes with a focus on monogenic forms; to identify and characterise additional subtypes of diabetes; and to identify variants contributing to susceptibility to type 2 diabetes. The YDX study is ongoing and recruits adult patients with diabetes onset below 45 years of age. The research nurses recruited patients from seven local hospital diabetes centres and multiple primary care centres from the Thames Valley in the United Kingdom. Most individuals with diabetes onset below 45 years of age (~5,500) received letters inviting them to take part in the study and on average 20% responded. By February 2016 this study included 1158 participants. Approximately 65% of those individuals were C-peptide positive and GADA negative (~660).

Whole blood and urine samples were collected from every patient and stored in -80°C. Plasma and serum samples were derived and stored. Clinical information on age of diabetes onset, treatment history, current medications and family history of diabetes were recorded and anthropometrical measurements collected. Basic biochemical tests such as HbA1c, fasting plasma glucose, renal function and lipid profile were measured in the recruitment site laboratories.

All patients signed an informed consent and ethical approval from the local ethics committee was obtained. The study was conducted in accordance with the Declaration of Helsinki.

Patient clinical characteristics in this study are included in Table 2.1.1. Participants from this study (n=581) were included in the biomarkers study (chapter 5), and also in the HNF1A-MODY (n=35) and type 2 diabetes (n=35) groups of urinary steroid metabolites study (chapter 6).

Study	Young Diabetes in Oxford (YDX) study	Protein fucosylation study (Croatian participants)
Number of patients	1158	477
Gender [number of men (%)]	689 (59.5)	268 (56.5%)
Age at recruitment [mean (SD) years]	45.74 (12.89)	47.43 (9.79)
Age at diagnosis [mean (SD) years]	30.54 (11.14)	35.22 (6.25)
BMI [median (IQR) kg/m ²]	29.51 (7.65)	29.40 (7.74)
FPG [median (IQR) mmol/L]	8.35 (4.7)	8.00 (3.40)
HbA1c [median (IQR) %]	7.95 (2.1)	7.30 (1.90)
Duration of DM [median (IQR) years]	12.15 (16.98)	11.00 (11.00)
C-peptide [median 9IQR) nmol/L]	0.44 (0.72)	0.72 (0.53)
Total cholesterol [mean (SD) mmol/L]	4.36 (1.19)	4.99 (1.22)
HDL [median (IQR) mmol/L]	1.20 (0.51)	1.20 (0.40)
TG [median (IQR) mmol/L]	1.20 (1.12)	1.60 (1.26)
Treatment		
Insulin +/- OHA [number of case (%)]	726 (62.7)	168 (38.8)
OHA [number of case (%)]	396 (34.2)	243 (50.9)
diet [number of case (%)]	36 (3.1)	66 (13.9)

Table 2.1.1 Patient clinical characteristics in YDX and Protein fucosylation studies. Variables with normal distribution presented as mean (SD), non-parametric distribution as median (IQR) and proportions as percentages; OHA, oral hypoglycaemic agents; BMI, body mass index; FPG, fasting plasma glucose.

2.1.2 Croatian participants in the Protein fucosylation in diagnosis of MODY study

This study was set up for the purpose of measuring N-glycan profile in adult subjects with diabetes diagnosed before 45 years of age and is described in chapter 5.

Initial selection of patients to the study was performed by diabetologists in Croatia using the Croatian national registry of patients with diabetes (CroDiab), which was established to determine the prevalence and incidence of diabetes and its complications at a national level. All primary and secondary care physicians looking after patients with diabetes were obliged to supply data to The Vuk Vrhovac University Clinic for Diabetes, Endocrinology and Metabolic Diseases on an annual basis until 2015.

The inclusion criteria for the study were diabetes onset before the age of 45 years and above 18 years at recruitment; evidence of endogenous insulin secretion (fasting or random C-peptide ≥ 0.2 nmol/L); and negative GAD (glutamic acid decarboxylase) autoantibodies. Patients were recruited by the Croatian diabetologists lead by Dr Edita Pape-Medvidovic at Vuk Vrhovac University Clinic for Diabetes, Endocrinology and Metabolic Diseases and University Hospital Merkur, Zagreb, in the period between June 2013 and June 2015. A blood sample was collected from each patient and stored in -80°C . Data collection included patient age, gender, age of diabetes onset, treatment history and current medication, family history of diabetes, weight, height, BMI and blood pressure. The biochemical data contained fasting plasma glucose, HbA1c, C-peptide, fasting lipid profile and creatinine. All recruited patients signed informed consent. The Ethics Committee of Clinical Hospital Merkur and the Faculty of Pharmacy and Biochemistry,

University of Zagreb approved the study. There were 13,035 adult patients with diabetes onset before the age of 45 years in CroDiab. The decision to measure GAD antibodies was based on a clinical assessment and around 33% of them had antibodies measured. This approach was likely to select more lean patients with T1D-like or MODY phenotype. The patients were prioritised if GAD antibody had been analysed at the diabetes centre, which participated in an antibody standardisation programme. From the list of eligible individuals, 2,929 patients throughout Croatia who fulfilled the inclusion criteria were invited to participate. Three hundred and twenty five patients responded and were recruited. A further 147 recently diagnosed patients were recruited from the Vuk Vrhovac University clinic. Additionally, 14 relatives were recruited for co-segregation studies. Of these, five individuals had diabetes, resulting in a total of 477 individuals with C-peptide positive non-autoimmune diabetes recruited in the study and all were included in the biomarkers study described in chapter 5. The clinical characteristics of those 477 patients are provided in Table 2.1.1.

2.1.3 Healthy subject group (Urinary steroids study)

The 35 healthy volunteers for the urinary steroid metabolites study (chapter 6) were selected from the 137 participants included in The Healthy Human Aging study, which were recruited by the research nurses from the local Birmingham population. This study involved people aged 20-81 years and BMI 20-30kg/m². These subjects had fasting plasma glucose measured to exclude undiagnosed diabetes and did not have history of ischemic heart disease, cardiovascular disease, respiratory disease or epilepsy. They were not

taking any steroids and were not on statin. Participants collected a 24 hours urine sample, which was used to measure their urinary steroid profile in the Institute of Metabolism and Systems Research, University of Birmingham and used as a control data set for subsequent urinary steroid studies.

2.2 Preparation of the plasma and serum samples and storage

Blood was taken and processed by research nurses at the recruitment sites. Plasma samples were obtained from whole blood collected into the vacuum green top tube with lithium heparin (BD Diagnostics, Franklin Lakes, NJ, USA) by centrifuging at 1000g for 10 minutes and stored at -80°C in apex tubes. Serum samples were collected into the vacuum gold top tube with the clot activator and left at room temperature for 30 minutes to form a clot, then centrifuged at 1000g for 10 minutes and stored at -80°C.

2.3 DNA extraction and quantification

Genomic DNA for *HNF1A* sequencing of all patients described in chapter 3 was extracted from the whole blood by Dr Carla Burrows in the Diabetes Research Laboratories in Oxford Centre for Diabetes, Endocrinology and Metabolism (OCDEM), University of Oxford and by Ms Tamara Pavic in the Zagreb Biochemistry and Molecular Biology Laboratory, University of Zagreb. Blood samples from the UK participants were obtained from the patients on the recruitment site and stored in -80°C until transferred to the Diabetes Research Laboratories in Oxford. DNA extraction in Oxford was performed on the automated Maxwell 16 system (Promega, Madison, WI, USA) using a Blood DNA

purification kit (Promega, Madison, WI, USA). The Maxwell system uses paramagnetic particles (MagnaCel particles), which provide mobile solid phase optimizing sample capture and purification. 300µL of the whole blood sample was briefly mixed with 30µL of Proteinase K and 300µL of Lysis buffer (both Promega, Madison, WI, USA), vortexed and incubated on the heat block at 56°C for 20 minutes. Subsequently, it was transferred to the well of the cartridge for automated extraction. In Zagreb laboratories DNA was extracted using a QIAamp DNA Purification kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol. DNA samples were stored in 1.4mL FluidX tubes at 4°C in both laboratories and Croatian samples were transported to the Oxford laboratory for *HNF1A* sequencing.

I quantified the genomic DNA using a NanoDrop ND-8000 spectrophotometer (ThermoScientific, Wilmington, DE, USA). The concentration of DNA was measured using the sample absorbance at a wavelength of 260nm. Before the first sample was applied, the spectrophotometer was blanked with 2 µL nuclease-free water, then the plasmid DNA concentration measured using 2µL sample.

2.4 DNA amplification by polymerase chain reaction (PCR)

I amplified the genomic DNA using polymerase chain reaction with forward and reverse primers for all 10 *HNF1A* exons, from all 1032 patients who had their *HNF1A* sequenced subsequently. I performed this work, the post-PCR product clean-up, gel electrophoresis and DNA sequencing reactions in the Diabetes Research Laboratories in OCDEM. Primers were ordered from Eurofins MWG Operon (Ebersberg, Germany) according to a

previously published sequence (Ellard, Bellanne-Chantelot et al. 2008) with an M13 tag attached at the 5'-end for later sequencing. The sequences of the primers are listed in chapter 3, Table 3.2.1. Primers were re-suspended to make 100pmol/ μ L solution, which was diluted in a 1:10 ratio.

Before PCR, stored DNA samples were transferred to the 96-well skirted plates (ThermoScientific, Cramlington, UK) using Freedom EVO platform (Tecan Group Ltd., Mannedorf, Switzerland), which allowed automated preparation of DNA concentration at 15ng/ μ L with nuclease-free water to 60mL total volume. All PCR reactions were performed in the 96-well skirted plates. 12.5 μ L PCR reaction included 2.5 μ L DNA (15ng/ μ L), DNA polymerase AmpliTaq Gold (5U/ μ L, Applied Biosystems, Foster City, CA, USA), 10x AmpliTaq buffer, dNTPs (Applied Biosystems), pre-prepared forward and reverse primer for the amplified exon of *HNF1A* (Eurofins MWG Operon, Ebersberg, Germany) and variable concentration of MgCl₂ depending on the exon being amplified (1.5mM solution for exon 1, 3, 4 and 10; 2mM for exon 2, 5-9). All reagent volumes and concentrations are provided in Table 2.4.1. Plates were sealed and centrifuged at 1000 RPM for 1 minute on a Rotina 420R centrifuge (Hettich, Massachusetts, USA). The thermocycling conditions using a PTC-225 Peltier Thermal Cycler (MJ Research, Waltham, MA, USA) are given in Table 2.4.1. Exon 9 thermocycling was extended by an extra 10 cycles to improve the quantity of post-PCR product. Plates with the post-PCR product were centrifuged at 1000 RPM for 2 minutes and stored at -20°C.

PCR reaction			
	1.5 mM MgCl ₂ (Exon 1,3,4,10)	2.0 mM MgCl ₂ (exon 2,5-9)	
Reagent	Volume [μL]	Volume [μL]	
10x PCR Buffer	1.00	1.00	
MgCl2 (25mM)	0.60	0.80	
dNTP (2.5mM each)	1.00	1.00	
AmpliTaq Gold (5U/μL)	0.04	0.04	
Distilled H ₂ O	6.96	6.76	
Primer F or R (5pM each)	0.40	0.40	
DNA (15ng/μL)	2.50	2.50	
Total volume	12.50	12.50	
Thermal cycling conditions			
Step	Number of cycles	Temperature [°C]	Time [minutes]
DNA denaturation	1	95	15
Primer annealing and DNA extension	35 (45 for exon 9)	94	1
		62	1
		72	1
Final DNA elongation	1	72	10

Table 2.4.1 Standard PCR reagents and volumes for 12.5μL reaction per 1 sample with thermal cycling conditions.

2.5 DNA gel electrophoresis and imaging

To confirm adequate amplification of exonic DNA, I performed gel electrophoresis for all *HNF1A* amplicons. A 2% agarose gel was made using 5g of agarose (Sigma-Aldrich, Gillingham, UK) in 250mL of 1x TBE (Tris, borate, ethylenodiaminetetraacetic acid) buffer (NBS Biologicals, Huntingdon, UK). First, the solution was heated in an 800W microwave for 2-3 minutes to dissolve the agarose, then cooled down to a temperature of around 60°C, when 5μL of ethidium bromide (Sigma-Aldrich, Gillingham, UK) was added, mixed and poured into electrophoresis gel former with three 32-well combs *in situ*. Electrophoresis of DNA amplicons was conducted by loading 5μL of post-PCR product with

5 μ L of orange G loading buffer (Sigma-Aldrich, Gillingham, UK) per sample with a multichannel pipette. 5 μ L of 100bp DNA ladder (New England Biolabs Ltd., Hitchin, UK) was used as a reference and loaded at the beginning of each lane of samples. 1x TBE was used as a running buffer in the electrophoresis tank. Electrophoresis was run at 110V for 30 minutes. The gels were visualised using Chemi-Doc MP and Image Lab software (Bio-Rad Laboratories, Watford, UK) and an example of the imaging of 96 samples (1 plate) is provided in Figure 2.5.1.

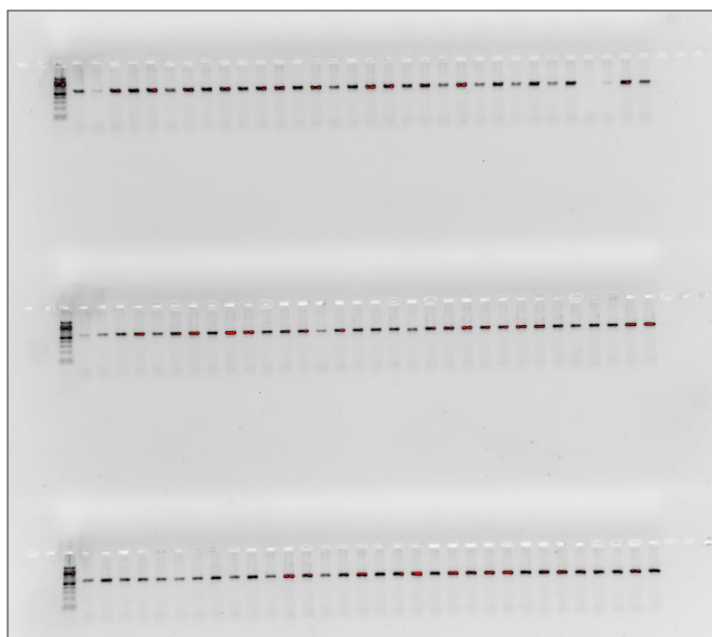


Figure 2.5.1 Example of a gel electrophoresis of 96 samples of post-PCR product (*HNF1A* exon 2).

2.6 DNA sequencing

I used bidirectional sequencing of all amplified *HNF1A* exons from 1032 patients recruited into biomarkers study using the Sanger sequencing method. The sequencing results are described in chapter 3.

2.6.1 Clean-up of PCR product

Exonuclease I (New England Biolabs, Hitchin, UK) and Thermosensitive Alkaline Phosphatase (Promega, Madison, WI, USA) in 1:1 volume ratio were used to remove unincorporated primers and dNTPs. 1 μ L of the mix was used per 2 μ L of post-PCR product per sample, centrifuged at 1000 RPM for 2 minutes and thermocycled at 37°C for 30 minutes, then heated to 80°C for 15 minutes to deactivate both enzymes and cooled down.

2.6.2 Sanger Sequencing

I used the modified bidirectional chain-terminating (Sanger) sequencing technique to sequence *HNF1A* amplicons. The technique is based on the selective incorporation of fluorescently labelled dideoxynucleotides (ddNTPs) by DNA polymerase during in vitro DNA replication (Sanger et al. 1977). ddNTPs are lacking 3'-hydroxyl group required for the formation of phosphodiester bond between nucleotides, therefore the DNA chain terminates after ddNTP is added. The concentration of labelled ddNTPs is around 100-fold lower than unlabelled dNTPs allowing for long enough DNA fragments to be produced. Finally, a DNA sequencer detects the different fluorescent dye label of ddNTPs during capillary electrophoresis.

I performed the sequencing reaction using Big Dye Terminator Sequencing mix v1.1 (Applied Biosystems, Birchwood, UK), Big Dye 5x dilution buffer, M13 sequencing primer (forward or reverse appropriately) and sterile water to the total of 7 μ L per 3 μ L of post-Exo I-TSAP treated sample (DNA template). The Big Dye Terminator mix included

fluorescently labelled ddNTPs, unlabelled dNTPs and DNA polymerase. M13 primers (forward 5'-TGTAACGACGGCCAGT-3', reverse 5'-CAGGAAACAGCTATGACC-3', Eurofins MWG, Operon) were compatible with the M13 tag on the *HNF1A* exon primers used in previous PCR reactions. Sequencing reaction reagents and thermocycling conditions are listed in Table 2.6.1. Post-sequencing products were briefly centrifuged, diluted with 5µL of distilled water per sample and underwent clean-up to remove unused reagents and incorporated primers/dNTPs by using 96-well Performa DTR Ultra plates (Edge BioSystems, Gaithersburg, USA). The post-sequencing products were passed through the Proforma DTR plate by centrifuging at 850g for 5 minutes. Eluted DNA samples were collected into 96-well semi-skirted plate (ThermoScientific, Cramlington, UK) and subsequently dried at 70°C for 90 minutes on a thermal cycler. The final sequencing reading was obtained using ABI 3730 DNA analyser (Applied Biosystems, Birchwood, UK) by Lynne Richardson team in the sequencing facility of the Department of Zoology, University of Oxford and the results were uploaded to the shared server.

Clean-up of the post-PCR product			
Reagent	Volume [μL]		
Thermosensitive alkaline phosphatase	0.5		
Exonuclease I	0.5		
Post-PCR product	2.0		
Thermal cycling conditions for the post-PCR product clean-up			
Step	Temperature [°C]		Time [minutes]
Enzymatic reaction	37		30
Deactivation of enzymes	80		15
DNA sequencing reaction			
Reagent	Volume [μL]		
Big Dye Terminator Sequencing Mix v1.1	0.250		
Big Dye 5x Dilution Buffer	1.875		
M13 Seq Primer F or R (2.5pM each)	2.000		
Distilled H ₂ O	2.875		
Post-PCR and post clean-up product	3.000		
Total volume	10.000		
Thermal cycling conditions for DNA sequencing reaction			
Step	Number of cycles	Temperature [°C]	Time
DNA denaturation	25	96	10 seconds
Primer annealing		50	5 seconds
DNA extension		60	4 minutes

Table 2.6.1 Reagents, volumes and thermal cycling conditions for post-PCR product clean-up and DNA sequencing reaction. Volumes given per 1 sample.

2.7 Genotyping

I performed genotyping of two common *HNF1A* variants (rs1169288, p.I27L; rs1800574, p.A98V) on the first four 96-well DNA sample plates, which had already undergone Sanger sequencing of *HNF1A*, to assess the sequencing quality with another method. This work was done in the Diabetes Research Laboratories in OCDEM.

2.7.1 TaqMan PCR

2µL of genomic DNA samples (15ng/µL) were transferred from four 96-well plates to one 384-well plate (4titude, Wottam, UK) using FluidX (Apricot Design Ltd., Covina, CA, USA) and dried at 80°C for 10 minutes on a thermal cycler. 3µL of PCR Master Mix (TaqMan, Applied BioSystems, Foster City, CA, USA) and 0.04µL of TaqMan SNP genotyping Assay for the examined SNP (Applied Biosystems, Foster City, CA, USA) was added to each well in the 384-well plate with dried DNA samples. TaqMan PCR Master Mix included Amplitaq Gold DNA polymerase, dNTPs and optimised buffer. TaqMan SNP genotyping Assay contained the genotyped SNP forward and reverse primers (at 900nM concentration) and 2 TaqMan MGB probes (one probe labelled with VIC dye for detecting allele 1 and the second probe labelled with 6FAM dye to detect allele 2, at 200nM concentration). The Plate and samples were briefly centrifuged and PCR performed on thermocycling conditions given in Table 2.7.1.

TaqMan PCR reaction			
Reagent	Volume [μL]		
Genotyping PCR Master Mix	1.50		
Genotyping assay (including primers)	0.04		
Distilled H ₂ O	1.50		
DNA (15ng/μL)	2.00 (dried before reaction)		
Total volume	3.04		
Thermal cycling conditions			
Step	Number of cycles	Temperature [°C]	Time
AmpliTaq Gold activation	1	95	15 minutes
DNA denaturation	40	92	15 seconds
Primer annealing and DNA extension		60	1 minute

Table 2.7.1 Genotyping reagent volumes per 1 sample and thermal cycling conditions for TaqMan PCR.

2.7.2 TaqMan genotyping and analysis

Post-PCR products were centrifuged at 1000g for 1 minute and TaqMan ABIPrism 7900HT Sequence Detection System (Applied Biosystems, Foster City, CA, USA) with SDS v2.3 software was used for allelic discrimination end-read. Results were visually checked to confirm correct assignment of allele calls.

2.8 Analysis of *HNF1A* variants

2.8.1 Mutation surveyor

All Sanger sequencing results were uploaded as ab.1 files and I analysed them using Mutation Surveyor DNA Variant Analysis software v5.0.1 (Soft Genetics, Cambridge, UK) and an example of the output is given in Figure 2.8.1. Human *HNF1A* sequence (ID 6927, chromosome 12, NC_000012.11; NM_000545.5; 121415861..121440315) from NIH

GenBank 105 annotation release [GRCh37.p13 (GCF_000001405.25)] was used as a reference file to compare and detect nucleotide substitutions. All results with sequencing quality below 50 and detected variants were visually checked to ensure correct interpretation of the software. Sequencing was repeated if the sequencing quality was below 30. Results were confirmed by re-sequencing from re-extracted DNA samples for all variants detected.

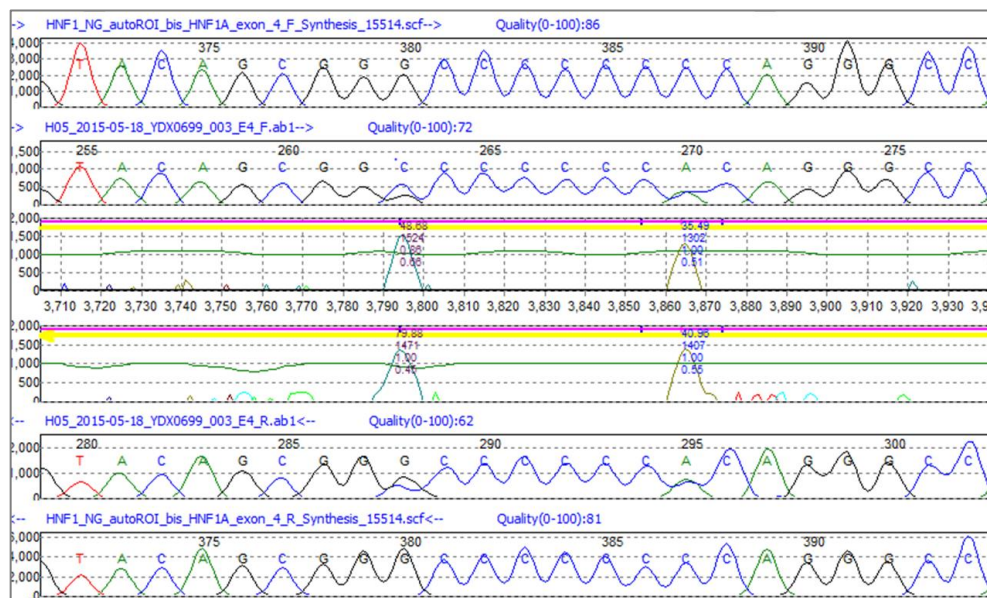


Figure 2.8.1 Example of the Mutation Surveyor output of the bidirectional sequencing (2 nucleotide substitutions captured: c.864G>C and c.871C>A).

2.8.2 Bioinformatics prediction tools

I analysed all non-synonymous *HNF1A* variants detected by Sanger sequencing for their predicted functional effect using 4 publically available bioinformatics prediction tools: Mutation Taster 2 (<http://mutationtaster.org/>) (Schwarz et al. 2014); Sorting Intolerant from Tolerant (SIFT, <http://sift.icvi.org>) (Kumar et al. 2009); Polymorphism Phenotyping

v2 (PolyPhen-2, <http://genetics.bwh.harvard.edu/pph2/index.shtml>) (Adzhubei et al. 2010); and Protein Variation Effect Analyzer (Provean, <http://provean.jcvi.org/index.php>) (Choi et al. 2012).

Mutation Taster 2 was used first, as it gives a very detailed descriptive output including the position of the mutation in the chromosome, previous reports of the variant, presence of the variant allele in sequencing databases such as 1000 Genomes and Exome Aggregation Consortium (ExAC), conservation data comparing amino acid (AA) sequences in Humans versus 10 other species, demonstrating evolutionary conservation, and lists protein features potentially affected by AA change. It gives a binary output of a variant to be “disease-causing” or “polymorphism”. The authors claim better prediction performance of this tool (C-statistics for detecting all known variants from 1000G and ClinVar of 90% comparing to around 86-87% for SIFT, PolyPhen-2 and Provean). Despite this, the tool is most discordant when all 4 are used together, which calls its accuracy in to question.

SIFT gives a prediction score (0-1, cut-off point 0.05) for non-synonymous variants based on the assumption that important AAs for protein function have been conserved throughout evolution. The SIFT prediction outcome is “damaging” or “tolerated”. A score of 0 means that AA in the examined position is highly conserved, predicting the variant to be damaging.

PolyPhen-2 takes into account 8 sequence-based and 3 protein structure-based features and gives 3 outputs: “probably damaging”, “possibly damaging” and “benign”. Provean

prediction is based on similar features and takes into account 58,000 human protein variants from UniProt with known functional outcomes. It gives a binary outcome, “deleterious” or “neutral”, with a prediction cut-off of -2.5. An effect on the splicing was examined using Human Splicing Finder (<http://www.umd.be/HSF3/index.html>) (Desmet et al. 2009) and Mutation Taster 2.

All protein truncating variants were assumed to be damaging to protein function and were predicted by Mutation Taster 2 as disease-causing.

2.9 Functional studies of HNF1A protein variants

I performed all steps of the functional assessment of HNF1A protein variants in the Diabetes Research Laboratories in OCDEM. A pipeline for the functional characterisation of HNF1A protein variants was already established in the Prof. Gloyn laboratory and I used the local protocols. The transcription activity was assessed using dual luciferase reporter system, protein expression by western blotting and DNA binding by electrophoretic mobility shift assay and results are presented and discussed in chapter 4.

2.9.1 Site-directed mutagenesis (SDM)

cDNA of human *HNF1A* (NM_000545.5) was inserted into the pcDNA 3.1/His C plasmid including ampicillin resistance sequence (*gift from KG Jebsen Center for Diabetes Research, University of Bergen; Figure 2.9.1*) and used as a template in mutagenesis of *HNF1A*. Primers to introduce *HNF1A* variants into the pcDNA 3.1/His C plasmid were designed using the Quikchange Primer Design website (Agilent Technologies, Berkshire,

UK). Primers were re-suspended to make 100pmol/ μ L solution, which was diluted 1:10. For 50 μ L SDM PCR reaction, 125ng of forward and 125ng of reverse primer were added per 50ng of pcDNA template. Table 2.9.1 lists all the reagents. *Pfu*Turbo DNA polymerase (2.5U/ μ L, Agilent Technologies, Berkshire, UK) was used to amplify the entire plasmid in the thermocycling reaction on PTC-225 Peltier Thermal Cycler (MJ Research, Minnesota, USA) with conditions described in Table 2.9.1. Subsequently, a restriction enzyme *Dnp1* (10U/ μ L, Agilant Technologies, Berkshire, UK) specific only to methylated DNA was used to digest non-mutagenized plasmid DNA at 37°C for 1 hour. 10 μ L of pre-digestion product could be run on 1% agarose gel at 80V to check if the SDM had worked. The SDM final product was stored at -20°C.

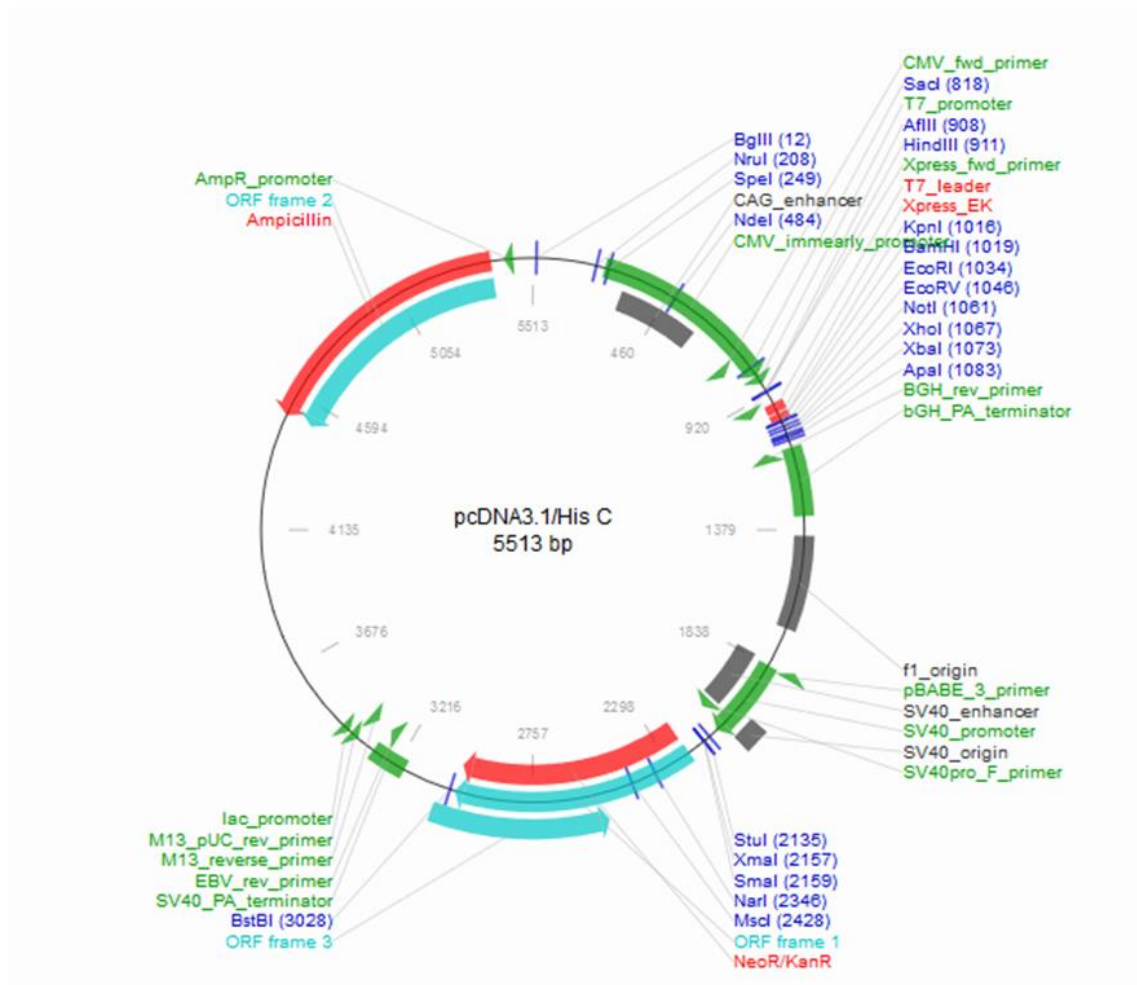


Figure 2.9.1 pcDNA 3.1/His C plasmid used as vector for HNF1A in the functional work. Image obtained from <https://addgene.org>.

SDM reaction			
Reagent	Volume [μL]		
10x buffer	5		
Specific forward primer (125nM)	as calculated for 125ng		
Specific reverse primer (125nM)	as calculated for 125ng		
<i>Pfu</i> Turbo DNA polymerase (2.5U/μL)	1		
dNTPs (2.5nM each)	1		
pcDNA <i>HNF1A</i> WT template	as calculated for 50ng		
Nuclease-free water	to the final volume of 50μL		
Thermal cycling conditions			
Step	Number of cycles	Temperature [°C]	Time [minutes]
DNA denaturation	1	95	2
Primer annealing	18	95	1
		55	1
		68	8

Table 2.9.1 The reagents, volumes and thermal cycling conditions for site-directed mutagenesis. Volumes given per SDM of 1 *HNF1A* variant.

2.9.2 Transformation of *Escherichia coli* and growth

E.coli strain was used to produce sufficient amounts of mutagenized plasmids before they were introduced to HeLa cells. XL1-blue *E.coli* cells were transformed with mutagenized plasmids (2 μ L of post-Dnp1 digestion PCR product per 40 μ L of defrosted competent *E.coli* cells), incubated for 30 minutes on ice and then exposed to 42 $^{\circ}$ C heat shock for 45 seconds to introduce plasmids into the bacteria. The *E.coli* cells were spread thinly on a pre-prepared sterile agar plate with ampicillin (100mg/mL) and grown at 37 $^{\circ}$ C for 18 hours. As the plasmids were carrying ampicillin resistance sequence, only transformed *E.coli* survived and multiplied. Two isolated colonies (A and B) from each plate were selected and cultured in 10ml of LB broth with ampicillin at 37 $^{\circ}$ C for 18 hours in a shaking incubator (311DS Labnet, Labnet International Inc., Woodbridge, NJ, USA). The next day,

500µL of the cell suspension in 500 µL of DMSO was used to create DMSO stock of each *HNF1A* variant transformed *E.coli* and stored in -80°C. The rest of the transformed *E.coli* suspension was pelleted by centrifuging at 4900g for 15 minutes (Heraeus Multifuge 3 S-R, Hanau, Germany), supernatant removed and pellet used for plasmid DNA extraction to make mini-prep.

2.9.3 Plasmid DNA extraction from transformed *Escherichia coli*

I extracted mutagenized plasmid DNA from harvested *E.coli* cells using the Pure Yield Plasmid Midi-prep System (Promega, Southampton, UK) according to the manufacturer's protocol. Briefly, the pellet of cells from 100mL bacterial culture was resuspended in 3mL of Cell Resuspension Solution, and 3mL of Cell Lysis Solution was added, mixed gently and incubated for 3 minutes at room temperature. Subsequently, 5mL of Neutralisation Solution was added and gently mixed by inverting. The lysate was centrifuged at 4900g for 25 minutes (faster if separation of debris and lysate not achieved) and supernatant was poured into the Pure Yield Clearing column to purify lysate after vacuum (Welch, Sheboygan, USA) applied. Plasmid DNA binds to the membrane of the underneath binding column. 5mL of Endotoxin Removal Wash was used next and vacuum applied, followed by 20mL of Column Wash Solution and vacuumed until dry (2-3minutes). To release DNA from membrane, a binding column was placed above a 1.5mL microcentrifuge tube in the Vacuum Elution Device. 600µL nuclease-free water was evenly applied to the membrane, left for 2 minutes prior to collection and passed through the membrane a second time to maximise DNA yield.

2.9.4 DNA quantification

Plasmid DNA was quantified using a NanoDrop ND-1000 spectrophotometer (ThermoScientific, Wilmington, DE, USA). The plasmid DNA concentration varied from 250ng/μL to 1100ng/μL.

2.9.5 Plasmid DNA sequencing

Extracted plasmid DNA with SDM modified *HNF1A* underwent Sanger sequencing using T7 forward primer and designed *HNF1A* cDNA forward primers F2, F3, F4, F6 (Eurofins Genomics, Wolverhampton, UK) to confirm the presence of the desired nucleotide change and ensure that no *HNF1A* additional nucleotide substitutions in the remaining part of the gene has been introduced. The sequences of the primers are listed in Table 2.9.2. For sequencing reaction 1.5μg of plasmid DNA, 2μL of primer (10pmol/μL) and nuclease-free water to 15μL were applied. The mixes were sent to Eurofins for sequencing according to their standard protocol.

Primer name	Forward (F) primer sequence (5'-3')
T7_F	TAATACGACTCACTATAGGG
HNF1A_F2	GTGGCGAAGATGGTCAAGTC
HNF1A_F3	AGGAGGAGCGAGAGACGCTA
HNF1A_F4	TAGTGACAGTGTCTACACC
HNF1A_F6	TTCACACGCCGGCATCTCAG

Table 2.9.2 PCR and M13 forward sequencing primers pcDNA 3.1 plasmid used to confirm desired mutagenesis outcome.

Once mutagenized plasmids with *HNF1A* variants were confirmed, DMSO stock of each previously stored variant was used to generate more plasmid pcDNA by growing

transformed *E.coli* in 100mL of LB broth with ampicillin for 21 hours to produce midi-preps after pcDNA extraction.

2.9.6 Maintenance of HeLa cell line and seeding technique

I used HeLa cell line for all functional work evaluating *HNF1A* variants, which is described in chapter 4. This cell line was originally derived from human cervical cancer cells (Scherer et al. 1953) and chosen due to a lack of endogenous *HNF1A*. HeLa cells were maintained in 1x Dulbecco's Modified Eagle's Medium (DMEM, Gibco, Thermo Fisher Scientific, Waltham, MA, USA) with 10% Foetal Calf Serum (FCS, Gibco, Thermo Fisher Scientific, Waltham, MA, USA) and 1% penicillin/streptomycin (10,000U/mL; Life Technologies, Paisley, UK). HeLa cells were split twice a week (1:8), maintained in T75 tissue culture flasks (Sarstedt, Leister, UK) and incubated at 37°C and 5% CO₂. Before splitting, cells were usually visually confluent at 90-98%. First, under a negative-pressure hood, the medium was removed and cells washed with 5mL of 1x phosphate buffered saline (PBS, Sigma-Aldrich, Gillingham, UK). Next, 1mL of TrypLE Express (Life Technologies, Carlsbad, CA, USA) was evenly spread on the cell surface to mobilise them and incubated at 37°C for around 6 minutes. The detachment of the cells was then confirmed under microscope, Nikon Eclipse TS100 (Nikon UK Ltd, Surrey, UK), diluted in 10mL of growth DMEM media and 1mL of mobilised cell solution was transferred to a new flask with 9mL of growth DMEM media pre-warmed to 37°C under a negative-pressure hood.

Cells were split 24 hours before transfection and counted using a cell counter (Cellometer Auto T4, Nexcelom Bioscience, Sarisbury Green, UK). HeLa cells were plated at a density

of 2×10^6 cells into 10cm² Petri dishes for future western blotting and EMSA and at 5×10^4 cells per well of the 24-well culture plates for transactivation assessment in DMEM media with 10% FCS, without antibiotics. The transfection was performed when cells were at 70-90% visual confluence.

2.9.8 Transfection of HeLa cells with plasmid DNA

For the later assessment of transcription activity, I transfected HeLa cells in the 24-well culture plates with 1µg of plasmid DNA mixes, which included pcDNA of WT *HNF1A* or *HNF1A* variant, the empty plasmid, pGL3 plasmid containing rat albumin promoter and Firefly luciferase, pRL-SV40 control vector containing *Renilla* luciferase (Promega, Southampton, UK) and nuclease-free water to a total volume of 70µL. The amounts of each plasmid pcDNA used are listed in Table 2.9.3. Lipofectamine 2000 (Invitrogen, Thermo Fisher Scientific, Waltham, MA, USA) in the ratio of 3µL: 1µg of DNA mix diluted in 330µL of 1x Opti-MEM I medium (Gibco, Thermo Fisher Scientific, Waltham, MA, USA) was used for the transfection of HeLa cells (100µL of the above mix/well) and incubated for 20 min at room temperature. Three technical replicates per variant were generated and incubated in DMEM media with 10% FCS for 24 hours at 37°C. The medium was changed 4-5 hours after transfection. Twenty hours later, the cells were harvested.

For the later protein quantification and DNA binding assessment, a ratio of 15 µL of Lipofectamine 2000 to 5 µg of plasmid cDNA was used for transfection of HeLa cells in 10 cm² Petri dish. Cells were incubated at 37°C and 5% CO₂ for 24 hours before lysis.

pcDNA plasmid	pcDNA amount [ng]
pcDNA <i>HNF1A</i> - WT or <i>HNF1A</i> - variant	240
pcDNA-empty plasmid	490
pGL3-RApromoter- FL reporter vector	250
pRL-SV40 control vector	10
Nuclease-free water	to the total volume of 70μL

Table 2.9.3 Components of pcDNA mix for transfection of HeLa cells for future dual reporter luciferase assay. Amounts of pcDNA given for 1 pcDNA *HNF1A* variant mix.

2.9.9 Transcription activity studies using dual luciferase reporter system

To evaluate the transcription activity of the five uncharacterised *HNF1A* variants, I used a dual luciferase reporter system with rat albumin promoter as a reporter construct. Firefly luciferase reporter (cloned into pGL3 vector) was utilised as a measure of transcription activity of the wild-type or mutagenized *HNF1A*. *Renilla* luciferase reporter worked as an internal control (previously inserted into pRL-SV40). pGL3 vector included rat albumin promoter sequence with *HNF1A* recognition motif. Both luciferase-carrying vectors were previously purchased from Invitrogen (Thermo Fisher Scientific, Waltham, MA, USA) and their DMSO stock stored in -80°C for future use. Firefly luciferase catalyses the oxidation of luciferin, which emits light at 560nm, while *Renilla* luciferase catalyses oxidation of coelenterazine and emits light at 480nm. Thus, the emitted light could be detected sequentially.

24 hours later the transfection medium was removed, then HeLa cells were washed with PBS and lysed with 100μL/well 1x passive cell lysis buffer (Promega, Southampton, UK) placed on a see-saw rocker SSL4 (Stuart, Stone, UK) for 25 minutes, then placed in -80°C for 1 hour. A thawed cell lysate was mixed thoroughly and 10μL of lysate/sample was

transferred to the 96-well white flat bottom plates (Corning, Appleton Woods, Birmingham, UK) and used for measuring transcription activity using the Dual-Luciferase Reporter Assay System kit (Promega, Madison, WI, USA). 50µL per well of substrate for Firefly luciferase (LARII) then 50µL of substrate for *Renilla* luciferase (“Stop and Glow”) were injected by the injectors of Multimode Plate Reader, EnSpire (Perkin Elmer, Waltham, USA) and luminescence was measured. Relative luciferase units for Firefly luciferase were normalised against *Renilla* luciferase units, means of three technical replicates per variant were calculated and then divided by the mean normalised luciferase units of the wild-type (WT) HNF1A protein. Each variant result was given as a percentage of the WT HNF1A. Each experiment included empty pcDNA, WT *HNF1A*, 2 positive controls (p.T260M, p.R203H), 1 synonymous variant (p.H179H) and 1 variant associated with an increased risk of T2D (p.E508K). Each experiment was repeated twice on further passages of HeLa cells to obtain three biological replicates.

2.9.10 Harvesting and lysing HeLa cells for Western blotting and EMSA

24 hours post-transfection, HeLa cells in 10cm² Petri dishes had media removed, then were washed with cold PBS, scraped while on ice and transferred into 1.5mL Eppendorf tubes. Subsequently, they were spun at 1,500 RPM at 4°C for 5minutes and resuspended in 150µL per tube of 1x passive lysis buffer (Promega, Southampton, UK) containing proteinase inhibitor cocktail (cOmplete, Sigma-Aldrich, Gillingham, UK) and lysed on ice for 25 minutes. Cell lysate was spun again at 13,000 RPM at 4°C for 20 minutes on an

Allegra 64R Centrifuge (Beckman Coulter, High Wycombe, UK) and supernatant snap-frozen at -20°C in aliquots to use in western blotting and DNA binding assessment later.

2.9.11 Quantification of the total protein using Bradford assay

Total protein in the lysates was quantified to enable calculation of the same concentration of protein for each HNF1A protein variant before western blotting and DNA binding was examined. Protein assay dye, 1x Bradford reagent (Bio-Rad Laboratories, Watford, UK) was utilised to quantify the total protein in the lysate by colorimetric assay using VersaMax microplate reader and Softmax Pro v5.2 (Molecular Devices, Sunnyvale, CA, USA). The Bradford assay utilises the binding of Coomassie Brilliant Blue dye G-250 to proteins. In acidic conditions, the dye remains in a red cationic form and absorbs light of wavelength 470nm. When the dye binds to protein, it converts to a stable unprotonated blue form with wavelength absorption of 595nm. Therefore, the absorbance at that wavelength allows estimation of protein concentration in the complex. Increasing concentration of Bovine Serum Albumin (BSA, Sigma-Aldrich, Gillingham, UK) 0-10µL (1mg/mL) with 0.1mM NaOH was used to produce a standard curve (Table 2.9.4 and a standard curve example in Figure 2.9.2). 2µL of each protein lysate was mixed with 38µL of 0.1mM NaOH. 1mL of 1x Bradford reagent was added to each sample including 7 samples for the standard curve, vortexed and incubated at room temperature for 10 minutes. 200µL of each sample in duplicate was transferred to 96-well spectrophotometer plate (Corning Life Sciences, Amsterdam, Netherlands) and measured on VersaMax.

Standard sample [μg]	Lysis buffer volume [μL]	1mg/mL BSA volume [μL]	0.1M NaOH [μL]	Diluted 1:5 Bradford reagent [μL]
0	2	0	38	1000
1	2	1	37	1000
2	2	2	36	1000
4	2	4	34	1000
6	2	6	32	1000
8	2	8	30	1000
10	2	10	28	1000

Table 2.9.4 Setting up the standard curve samples using bovine serum albumin (BSA) and HeLa cell lysis buffer.

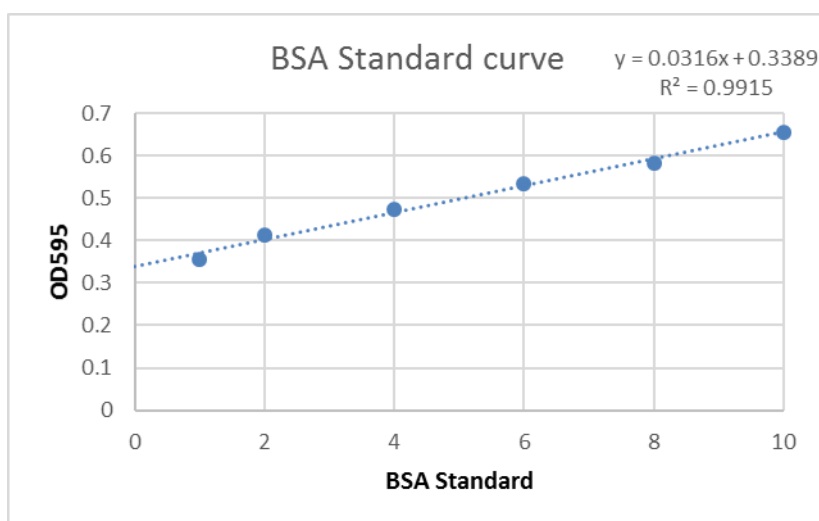


Figure 2.9.2 A BSA standard curve example.

2.9.12 Protein expression using Western blot analysis

I assessed protein expression by chemiluminescent western blotting, utilising anti-HNF1A antibodies to identify HNF1A. β -tubulin was used as an internal control.

4 μL of Laemmli loading buffer (Bio-Rad, Hemel Hempstead, UK) and 1 μL dithiothreitol (DTT, protein stabiliser, Sigma-Aldrich, Gillingham, UK) were added to 10 μg of total

protein per variant and denatured at 80°C for 20 minutes, and then loaded onto a 4-20% Criterion TGX Stain-free Precast gel (1mm, 26 wells, Bio-Rad, Hercules, USA). The Precision Plus Protein All Blue standards (Bio-Rad, Munich, Germany) was applied as a reference protein ladder. Electrophoresis was performed at 250V for 30 minutes. The gel was activated using the Chemi-Doc MP Imaging System (Bio-Rad, Watford, UK) and proteins transferred to a PVDF Midi membrane using Trans-Blot Turbo Midi PVDF Transfer pack (Bio-Rad, Watford, UK). Subsequently, the blot was cut to separate HNF1A protein from β -tubulin and blocked first with 5% milk in 1x tris-buffered saline tween (TBST) for at least 1 hour, then overnight with primary antibodies in 5% milk at 4°C (anti-HNF1A sheep polyclonal antibody 1:1000 and anti- β -tubulin rabbit antibody 1:2000, Santa Cruz Biotechnology, USA).

On day 2, primary antibodies were washed off with 1x TBST buffer and secondary antibodies in 5% milk (rabbit-anti-sheep and goat-anti-rabbit antibody respectively, 1:1250 both and conjugated with horseradish peroxidase, HRP, ThermoFisher Scientific) were applied for at least 1 hour at room temperature and washed off afterwards. The blot was exposed to Clarity Western ECL (Bio-Rad, Watford, UK), which contained HRP substrate and imaged with Chemi-Doc MP Imaging System. An example of western blot image is provided in chapter 4.

Densitometry of the protein bands was performed using Image Lab software (Bio-Rad, Munich, Germany), divided by the value calculated for β -tubulin band in each lane and

expressed as a percentage of the WT band. Finally, the mean of the percentage from 3 biological replicates for each HNF1A variant was calculated.

2.9.13 Assessment of DNA binding by Electrophoretic Mobility Shift Assay

I utilised the Electrophoretic Mobility Shift Assay (EMSA) to assess binding of the WT or variant HNF1A protein to a fluoro-labelled double-stranded oligonucleotide of the rat albumin promoter. EMSA is used to study protein-DNA complex interactions and is based on the principle that bigger protein-DNA complexes travel slower than unbound DNA in non-denaturing gel during electrophoresis, creating a “gel shift”. Adding an antibody against the examined protein (anti-HNF1A) forms an even bigger complex of antibody-protein-DNA and is visualised as a “supershift”. Both strands of the rat albumin oligonucleotide (5'-ACCAATAACTGTAGATCATTAACCACA-3', 5'-TGTGGTTAATGATCTACAGTTATTGGT-3', Eurofins Genomics, Wolverhampton, UK) were 5'-end-labelled with a Cy5 probe. Cy5 (indodicarbocyanine) is a fluorescent dye, which stabilises double stranded DNA and can be visualised with a near-infrared filter on the Chemi-Doc MP Imaging System. Forward and reverse Cy5-labeled oligonucleotides of the rat albumin promoter (100µM each) were mixed and annealed by heating to 100°C for 5 minutes, and then left to cool down slowly. They were then diluted 1:1000 (50nM) and stored in aliquots for future use.

On the day of running electrophoresis, 10 µg of a total protein of the WT or variant HNF1A was mixed with 7 µL of the binding mix (Odyssey EMSA kit, LiCor, Lincoln, USA), incubated for 20 min with 1µL of the Cy5-labelled oligonucleotide probe and loaded onto the 5%

Criterion Precast gel (Bio-Rad, Hercules, USA). Table 2.9.5 includes all reagents of the binding mix. 5 μ L of the anti-HNF1A antibody was added to the second sample of WT HNF1A to confirm the position of HNF1A band using the supershift effect and incubated for 30 min before the Cy5-labeled probe was added. Cy5 probe is light sensitive, therefore from this point, samples were covered with aluminium foil and I added loading dye and loaded samples to the gel in the room with very limited light exposure. The electrophoresis was performed at 100V over 1 hour and gel imaged using Chemi-Doc MP Imaging System, of which example is provided in chapter 4.

The bands were quantified by densitometry utilising Image Lab software and presented as a percentage of the WT. The experiment was repeated on subsequent passages of HeLa cells to obtain three biological replicates and means from all biological replicates were calculated for each HNF1A variant.

Reagent	Volume [μ L]
Volume for 10 μ g of total protein	as calculated
10x binding buffer	2
DTT/TWEEN 20	2
dI-dC	1
NP-40	1
MgCl ₂	1
Cy5-labelled probe	1
Nuclease-free water	to the total of 20 μ L

Table 2.9.5 Reagents included in the protein binding mix for Electrophoretic Mobility Shift Assay; NP-40, detergent; dI-dC, template for RNA polymerase (to reduce non-specific interactions). Volumes given for 1 sample.

2.10 Measuring C-peptide, GAD antibodies and CRP

2.10.1 C-peptide and high-sensitivity CRP assays

C-peptide from the UK participants and CRP assays were performed by Dr Neha Shah in the Biochemistry Laboratory of the Oxford University Hospitals NHS Trust. C-peptide from the Croatian participants was measured by the staff in the biochemistry laboratories of the recruiting hospitals in Croatia. In Oxford, C-peptide was analysed using a chemiluminescence immunoassay on an ADVIA Centaur analyser (Siemens Healthcare Diagnostics Ltd). The vast majority of UK samples (n=554) had their CRP measured using a wide range latex-enhanced immunoturbidimetric high-sensitivity assay on an ADVIA 2400 analyser (Siemens Healthcare Diagnostics Ltd) with a quantification limit of 0.01 mg/L. All Croatian samples and remaining samples from the UK (n=29) had CRP measured with a less sensitive method and the lowest quantifiable level of 0.1 mg/L (both methods are reproducible but demonstrate small accuracy difference at a level <1 mg/L (Shah et al. 2014).

2.10.2 GAD antibodies

Glutamic acid decarboxylase antibodies (GADA) were measured by radioimmunoassay using ³⁵S labelled GAD65 by Kyla Chandler in the Immunology Laboratories in the University of Bristol for patients recruited from the UK and by the immunology laboratories staff in the Croatian hospitals for participants recruited from Croatia. The cut-off for the positive result in the Bristol University laboratory was 13 WHO Units/mL initially

(samples measured n=218) and changed to 33 DK Units/mL later in the study following the National Institute of Diabetes and Digestive and Kidney diseases Harmonisation programme to standardise GAD antibody assays across the UK (Bonifacio et al. 2010). In Croatia, GADA were measured by ELISA immunoassay (EUROIMMUN AG). The cut off for the positive result in Croatia was 5 WHO Units/mL and the laboratory similarly participated in the Diabetes Antibody Standardization Program.

2.11 Release and labelling of N-glycans

All N-glycan profiling was done by Ms Tamara Pavic in the Laboratories of the Faculty of Pharmacy and Biochemistry, University of Zagreb and I assisted her in analysing one of the 5 batches of samples. Briefly, plasma N-glycans were first released from proteins, labelled with fluorescent 2-aminobenzamine to enable later detection and separated on liquid-liquid phase during chromatography. The results of plasma glycan analysis are discussed in chapter 5.

2.11.1 Randomisation to correct for the batch effect

Randomisation to correct for the batch effect was performed first to exclude impact of the procedure on results during a particular run of the analysis. For that reason, samples from different plates were mixed and allocated to the new 96-well plates including 10% of duplicates. All samples, which have already been sequenced and known to have the *HNF1A* variant were included in duplicates. Additionally, 4-5 standard plasma samples from the previous analysis with validated results and 1 blank sample were added per 96-

well plate. The correlation of duplicates was analysed afterwards and if not present, the outlying plate was run again.

2.11.2 Sample preparation

Plasma samples were denaturated with 2% SDS (sodium dodecyl sulphate, Invitrogen, USA) at 65°C for 10 minutes. 4% Igepal CA630 (Sigma Aldrich, USA) was added to overcome the effect of SDS on the enzyme and shaken on the plate shaker for 15 minutes. Plasma N-glycans were released using 1.2 Units of peptide N-glycosidase F (PNGase, Promega, Madison, WI, USA) in 10 µl of 5x PBS, which was added to each sample and incubated in 37 °C for 18 hours.

2.11.3 2-aminobenzamide labelling

A fluorescent tag, 2-aminobenzamide (2-AB, Sigma-Aldrich, St. Luis, MO, USA) solution, was used to label released N-glycans for later detection. Subsequently, 25 µl of solution per sample was added and incubated for 2 hours at 65°C and cooled down in room temperature. The 0.2 µm pre-conditioned GHP plate (Pall Corporation, Port Washington, NY, USA) containing a hydrophilic membrane was employed to capture glycans.

2.11.4 Clean-up of 2-AB labelled N-glycans and elution

After incubation, unbound labels and reducing agent were removed from the samples using hydrophilic interaction liquid chromatography solid-phase extraction (HILIC-SPE).

The 96-well PCR plate was used for collection of 2-AB labelled N-glycans by placing it under the GHP plate. 90µl of ultra-pure water was added to each well, shaken for 15 minutes and plates centrifuged at 1000 RPM for 5 min to collect the elute. Plates with eluted 2-AB-labelled N-glycans was sealed and stored at -20°C until analysis.

2.12 HILIC-ultra performance liquid chromatography of N-glycans

Fluorescently labelled *N*-glycans were separated by hydrophilic interaction chromatography (HILIC) on a Waters Acquity UPLC instrument (Waters, Milford, MA, USA) with a fluorescence (FLR) detector set with excitation and emission wavelengths of 330 nm and 420 nm respectively. The instrument was controlled through Empower 2 software, build 2145 (Waters, Milford, MA, USA).

The principle of HILIC is that the mobile phase forms a thin layer of water on the stationary phase creating a water-rich versus water-deficient acetonitrile (ACN) gradient and liquid-liquid extraction system.

The mobile-liquid phase inside the silica column was made by 100mM ammonium formate (pH=4.4) as solvent A (to control pH and ion strength of the mobile phase) and ACN as solvent B. Samples were maintained at 5°C and brought to 75% ACN before injection to replicate the conditions in the column. After the system had been equilibrated, the first blank sample (ultra-pure water) was injected, then standard 1 (2-AB-labelled dextran ladder), standard 2 (2-AB-labelled plasma pool used as reference), then 16 study samples. Both standards were then repeated. The separation method used

a linear gradient of 75–53% acetonitrile at a flow rate of 0.56 ml/min in a 25-minute analytical run at 25°C. Glycans eluted in an increasing order of polarity: the first eluted were small, neutral glycans and the biggest with the most charge i.e. containing sialic acids, last.

2.13 Quality control of plasma pool samples & standard 2 (dextran)

For dextran (2 samples), it was checked if 2 standard glycan peaks overlap.

For the previous patient plasma glycan pool samples (n=2), the area under the peak was compared for the biggest peaks 15 & 23 (Figure 2.13.1). Peaks should be symmetrical and not shifted. Study samples were checked for any peak anomalies such as split, new peaks or very small peaks compared to standard pool samples.

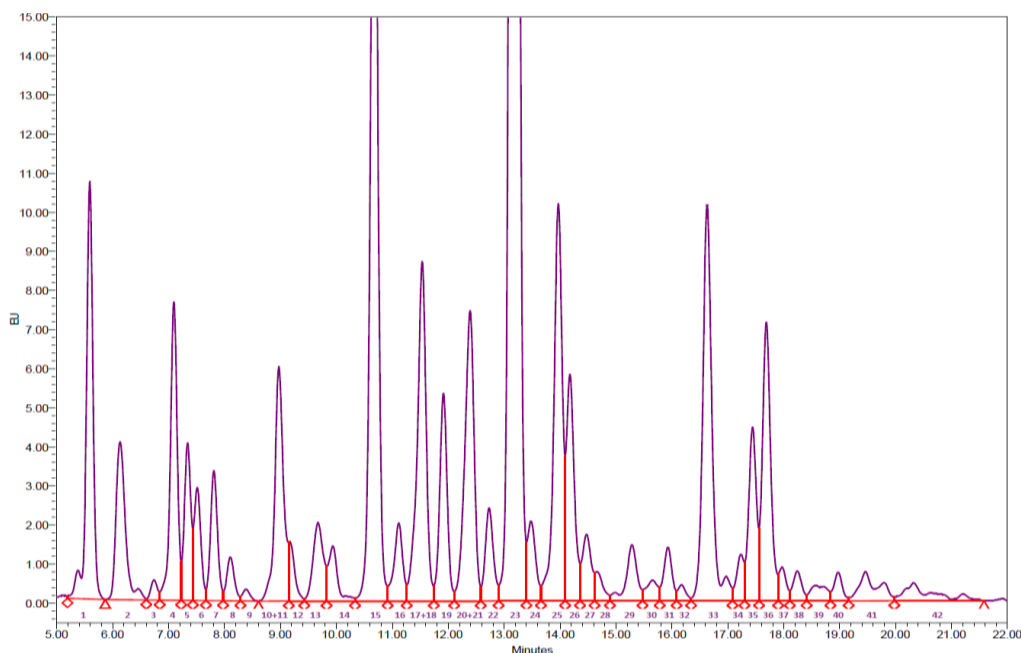


Figure 2.13.1 Chromatogram of the N-glycans of the previous patient pool sample used for QC. The tallest peak GP15 and 23 as marked above x axis; EU, emission units. Image shared by Tamara Pavic from University of Zagreb.

2.14 Integration of the chromatography peaks

Data processing was performed using a traditional integration algorithm after which each chromatogram was manually corrected to maintain the same intervals of integration for all the samples. The chromatograms were all separated in the same manner into 38 chromatographic peaks that enabled reliable quantification. The middle points of the baseline chromatogram between consecutive glycan peaks were marked, so the area under the curve (AUC) can be measured using rhomboid markers (as shown in Figure 2.13.1). The peaks were assigned according to Saldova et al. (Saldova et al. 2014).

The AUC for each GP peak represents the percentage of each glycan from the total N-glycans and was statistically analysed as such.

2.15 Gas chromatography mass spectrometry of urinary steroids

Gas chromatography mass spectrometry (GCMS) was used to evaluate the urinary steroid metabolite profile in patients with HNF1A-MODY, type 2 diabetes and healthy controls, which is presented and discussed in chapter 6. All samples were analysed in the Laboratory of the Institute of Metabolism and Systems Research, University of Birmingham by Beverly Hughes and Lorna Gilligan.

Thirty-three different urinary steroids and their metabolites were separated by gas chromatography/mass spectrometry on the Agilent 5975 instrument (Agilent Technologies Inc., Santa Clara, CA, USA) by the method described before (Shackleton 1986, Krone et al. 2010). In summary, steroid sulphate and glucuronide conjugates were first extracted from urine by solid phase extraction (SPE) and chemically released from conjugates by enzymatic hydrolysis to make them stable at high temperature. Freed steroids were re-extracted by SPE and derivatised to form methyloxime-trimethylsilyl ethers (first carbonyl groups were converted to methyloximes, then hydroxyl groups were silylated). The last stage was liquid-liquid extraction. Samples were injected sequentially and separated over 35 minutes on an increasing temperature gradient (100-300°C). The individual steroids were ionised in the fused silica column and specific ions formed from different metabolites. Internal standards, sigmasterol, cholesteryl butyrate and 3 β 5 β TH-Aldosterone were used for quantification and previously stored urine samples of healthy individuals were used for quality control. The values initially reported as $\mu\text{g/mL}$ were multiplied to give the final values per litre of urine.

2.16 Statistical analysis

I used IBM SPSS version 23 for the analysis of the results presented in Chapters 4-6. Medians with interquartile ranges (IQR) were calculated for non-parametric continuous variables and percentages for categorical variables. Non-parametric statistics (Mann-Whitney test) was used for comparison among two groups and Kruskal-Wallis test for comparing more than 2 groups. Differences of frequencies for categorical variables were tested using the Chi-squared or Fisher's exact test. Significant differences were generally assumed for $p < 0.05$, and were adjusted for the false discovery rate using the Benjamini-Hochberg procedure when the number of tests exceeded 10 (e.g. N-glycans or urinary steroid metabolites analysis). The prediction model for HNF1A-MODY was derived using binary logistic regression, where only variables correlating significantly with the binary outcome were included.

Results from each step of functional studies were presented as mean (+SEM, standard error of means) percentage of the WT HNF1A protein and compared using one-way analysis of variance (ANOVA) with pairwise comparison of mean percentage (+SEM) of the WT HNF1A for the investigated variant and p value was adjusted for multiple comparisons.

Chapter 3

***HNF1A* sequencing from the subjects with young adult onset non-autoimmune diabetes and interpretation of the *HNF1A* variants**

A manuscript including some of the content of this chapter has been published:

- Pavić T, Juszczak A, Pape Medvidović E, Burrows C, Bennett AJ, Šekerija M, Čuća-Knežević J, Gloyn AL, Lauc G, McCarthy MI, Gornik O, Owen KR. Maturity Onset Diabetes of the Young due to *HNF1A* variants in Croatia. *Biochem Med (Zagreb)*. 2018 Jun 15; 28(2):020703.

3.1 Introduction

The increase of large scale sequencing projects has resulted in increased genetic testing in clinical and research settings and this continues to identify novel variant alleles in Mendelian and common diseases. However, the interpretation of variant alleles remains challenging, especially in the case of very rare variants.

3.1.1 Assigning disease causality to the genetic variants

The guidelines for the interpretation of sequence variance in human disease were briefly discussed in chapter 1, section 1.5. In summary, the working group of experts in genomic research and clinical diagnostic sequencing convened by US National Human Genome Research Institute recommended the use of genetic, bioinformatic and experimental data if available; taking advantage of public data sets of genomic variants such as ExAC/GnomAD; and not regarding previous reports as definitive (MacArthur, Manolio et al. 2014). The group emphasised that all listed aspects of the interpretation of genetic variation should be analysed together and none of them singly offer sufficient evidence to assign disease causality to the allele (MacArthur, Manolio et al. 2014, Quintans, Ordonez-Ugalde et al. 2014, Kalia et al. 2017).

Below, I discuss all mentioned aspects of the guidelines in context of HNF1A-MODY to provide background for the assessment of rare nonsynonymous *HNF1A* alleles found in my study.

3.1.1.1 Interpretation of genetic results in association with clinical features

Pre-genome sequencing, diagnosis of the genetic disease was based mainly on the patient's clinical features and the known inheritance pattern of the disease.

As discussed in chapter 1, the typical phenotype of a HNF1A-MODY patient includes BMI in the normal range ($<25\text{kg/m}^2$), less than 30 years of age at presentation with diabetes, at least two generations' family history of diabetes, preservation of endogenous insulin secretion, negative β -cell antibodies and sensitivity to low doses of sulphonylureas. Those are strict criteria which would miss subjects with older presentation of diabetes or higher BMI. Therefore, for the purpose of my thesis, I defined HNF1A-MODY as diabetes with negative β -cell antibodies and C-peptide in the normal range in the individual with *HNF1A* allele co-segregating with diabetes. Sensitivity to small doses of sulphonylureas will further support the diagnosis of HNF1A-MODY, if available. This definition of HNF1A-MODY accepts some overlap with subjects with T2D.

Typical features of T2D are those present in associated metabolic syndrome such as central adiposity, elevated triglycerides, fatty liver, insulin resistance (expressed by acanthosis nigricans and C-peptide in the upper normal range) and BMI above 25kg/m^2 .

However, it has to be acknowledged that in reality there is a phenotypical spectrum of young adult-onset non-autoimmune diabetes and many such individuals will have a phenotype between MODY and T2D. The final phenotypic outcome is influenced not only by genetic factors, such as *HNF1A* alleles, but also by changes in multiple pathophysiological pathways, interactions between genes/variants and environmental

factors. This was recently described as a “palette model” of diabetes (McCarthy 2017). In the case of monogenic diabetes (e.g. HNF1A-MODY) the genetic component (*HNF1A* allele) plays a major role in the development of diabetes in those individuals.

To complicate the picture further, *HNF1A* alleles have a variable penetrance so individuals carrying the same *HNF1A* allele can present with diabetes before or after 30 years of age or remain non-diabetic throughout their life (Flannick, Beer et al. 2013). *HNF1A* variants with low penetrance can carry an increased risk of type 2 diabetes, such that the subjects’ phenotype may resemble more T2D.

Co-segregation of a particular *HNF1A* allele with diabetes in the family (and at the same time absence of the allele in non-diabetic subjects) supports the disease causality of the allele. On the other hand, lack of co-segregation with diabetes or an allele present in older non-diabetic subjects may suggest a neutral effect of the allele or low penetrance. Co-segregation is also less useful in the case of reduced penetrance of the variant (Quintans, Ordonez-Ugalde et al. 2014). Finally, *de novo* mutations occur with different frequencies in different genes and must be considered when co-segregation of the allele with diabetes is missing but there is strong clinical suspicion of monogenic diabetes.

3.1.1.2 Genetic evidence including allele frequency in large-scale sequencing samples

As deriving clinical evidence for the disease causality of an allele remains a challenge, genetics can provide further clues. A protein truncating *HNF1A* variant in the coding region or a splice site variant present in all isoforms of the protein (first 7 exons), and functionally essential part of the protein (such as a DNA binding domain), is more likely to

be damaging. However, some transcription errors may be overcome by alternative RNA splicing (MacArthur et al. 2012).

Large-scale collaborative exome/genome sequencing projects, such as GnomAD (described in section 1.5 of chapter 1) provide the largest dataset of human genetic variation, and characterise minor allele frequency (MAF) depending on ancestry. Generally, an MAF which is higher than expected for Mendelian disease (based on the prevalence of that disease) supports the neutral functional effect of the allele in a rare disease, such as HNF1A-MODY. At the same time, very low allele frequency does not automatically make the variant disease-causing. Each exome in ExAC contained an average of 7.6 rare non-synonymous variants (MAF<0.1%) in well-characterised Mendelian disease genes, of which clearly not all cause penetrant autosomal or recessive disease (Lek, Karczewski et al. 2016). The limitation of the ExAC/GnomAD database is that it includes some individuals with T2D, and therefore analysing allele frequency in a subgroups of ExAC with and without diabetes (T2D-GENES project) is also useful despite smaller number of participants (<http://www.type2diabetesgenetics.org>).

3.1.1.3 Bioinformatic prediction of the functional consequences of alleles

Multiple bioinformatic tools are available to guide the prediction of allele functional effects. These take into account genetic features of the variant such as evolutionary conservation (e.g. SIFT) and/or biochemical properties of amino acids, 3D protein structure and function (e.g. Polyphen 2, Provean). All bioinformatic tools are limited by

the extent to which a given gene/protein has been previously investigated and need continuous updating to be accurate.

In my study, I have used four publically available *in silico* prediction tools: SIFT, Polyphen2, Provean and Mutation Taster 2 which I described in detail in chapter 2, section 2.8.2 and in section 3.2.6 in this chapter.

3.1.1.4 Contribution of experimental data

Functional characterisation of the protein variant can contribute to the interpretation of its phenotypic effect. Human cell lines such as HeLa or INS1 can be helpful in assessing the effect of *HNF1A* alleles. The loss of HNF1A protein function in the experimental model can support the disease causality of the *HNF1A* allele (Yamada et al. 1999), on the contrary, protein function similar to the wild-type protein may suggest neutral effect of the *HNF1A* allele (Galan et al. 2011). However, a particular experimental model would not address all aspects of protein function so negative result is less useful.

Nonetheless, human biological pathways are complex and the results from an experimental model should be extrapolated to human disease with caution. An example is the *Hnf1a* knock-out mice, which are not diabetic in the heterozygous state but only as homozygotes, while in humans MODY is caused by heterozygous inactivating *HNF1A* alleles.

3.1.1.5 Previous reports in the literature or disease databases

Databases of known disease-causing variants such as the Human Gene Mutation Database (HGMD) or NCBI OMIM are useful provided they include all aspects of disease causality of the alleles discussed above and are regularly updated to account for new evidence. The 1000 Genomes project and NHLBI Exome Sequencing Project (ESP6500, n=6500) were initially the biggest datasets of human genome variation but had a relatively small sample size and were therefore not able to provide an accurate MAF for very rare variants. Also, rare alleles were more likely to receive a “damaging” label and were included in HGMD. An example of this is the p.G31D variant in *HNF1A* which in 2013 was considered MODY-causing (HGMD) while now it is not, as more evidence is available that this allele is neutral to the protein function.

3.1.2 Known rare alleles causing HNF1A-MODY and common *HNF1A* alleles

In 2013, the Department of Molecular Genetics from University of Exeter and Department of Genetics from University of Pierre and Marie Curie in Paris summarised all *HNF1A* variants found to-date in sequencing subjects clinically suspected to have MODY (Colclough, Bellanne-Chantelot et al. 2013). Alleles had been divided into disease-causing, common variants or variants of unknown significance (VUS) based on available evidence. This included co-segregation of the *HNF1A* allele with diabetes, bioinformatic prediction, results of the functional work (mainly transcription activity and some DNA binding of HNF1A mutant protein), MAF in ESP6500 and 1000 Genomes projects. The authors had reported 414 different disease-causing *HNF1A* alleles in 1247 families and 114 common

variants and variants of unknown significance in coding and non-coding regions of *HNF1A* (listed together without differentiation between common variants and VUS).

Among *HNF1A* variants reported by Colclough *et al.*, missense variants accounted for 54.7% and truncating variants for 31.1% (frameshift 21.7% and nonsense 9.4%). The splice site variants accounted for 8.7%, promoter region variants 1.9%, in-frame indels 2.4% and partial or whole exon or gene deletion 1.2% of *HNF1A* variants seen in those cases (Colclough, Bellanne-Chantelot *et al.* 2013). Exons 2 and 4 were the most commonly affected (0.33 and 0.28 variants per nucleotide), and exon 5 and 10 the least affected (0.11 and 0.07 variants per nucleotide). Most of the disease-causing *HNF1A* variants were reported in the dimerization and DNA binding domains (0.31 and 0.29 variants per nucleotide respectively), while fewer variants were reported in the transactivation domain and the *HNF1A* promoter (0.14 and 0.03 variants per nucleotide). The missense variants were more common in dimerization and DNA binding domains than in the transactivation domain (69% versus 48%).

It had been also observed that subjects with missense variants in exons 1-7 had an earlier diabetes onset than those with *HNF1A* variants in exons 8-10 [18 (12.5-23.5) versus 29 (19.5-38.5) years] (Harries, Ellard *et al.* 2006, Bellanne-Chantelot *et al.* 2008). This genotype/phenotype relation could be explained by the fact that alteration in exons 1-7 affect all three isoforms of *HNF1A*, while variants in exons 8-10 affect only isoform A. This leaves isoform B, which is most abundant in human islets, unchanged. This difference had not been reported in individuals with protein truncating variants, which all presented with

diabetes at younger age (18 and 20 years). This would be explained by nonsense-mediated mRNA decay (NMD). NMD prevents translation of mRNA transcripts with stop codons by eliminating them. This is difficult to explain as stop codons in exons 8-10 would not affect protein isoform B and C. The *HNF1A* gene is loss-of-function intolerant with a probability of Loss-of-function Intolerant (pLI) score of 0.97 with the cut-off for LoF intolerant genes of ≥ 0.9 (Lek, Karczewski et al. 2016).

The single most commonly reported variant allele in *HNF1A* is c.872dupC in exon 4 (p.G292fs, previous p.P291insC), which by 2013 had been reported in at least 234 families (Colclough, Bellanne-Chantelot et al. 2013). Another commonly affected codon after P291 was P379 in exon 6, which had 5 different missense and 6 frameshift variants reported at this position. Both codons P291 and P379 are in a poly-C region of the gene, and are more prone to C deletion or duplication.

In 2016, ExAC reported a total 802 variant alleles in *HNF1A* including rare and common variation (Lek, Karczewski et al. 2016), while to-date GnomAD reports a total of 1434 variants in *HNF1A*. This does not include most of the MODY-causing alleles which were individually too rare to be seen in population samples. The MODY-specific database of *HNF1A* variants published by Colclough *et al.* included variants collected by genetic laboratories sequencing *HNF1A* from subjects with high clinical probability of MODY (Colclough, Bellanne-Chantelot et al. 2013). ExAC/GnomAD combine exome/genome sequencing data of individuals mainly without diabetes in the large-scale sequencing samples and report more variants due to increased sample size. There were 4 amino-acid

(AA) substitutions reported in GnomAD in codon P291 (S, T, R and A in 62, 2, 8 and 1 case in ~ 235,000 alleles) and 4 AA substitutions in codon P379 (S, A, T and H in 9, 52, 9 and 12 cases in ~ 245,000 alleles). Most of them are consistent with the known prevalence of MODY with the exception of p.P291S and p.P379A which are more common.

The particular *HNF1A* variants found in my study are discussed in detail in sections 3.3-4.

3.1.3 *HNF1A* variants implicated in an increased risk of T2D

Multiple genome-wide association studies (GWAS) have identified more than 100 loci associated with T2D (Voight et al. 2010, Morris et al. 2012, Scott et al. 2017). Those include loci in and near monogenic diabetes genes such as *HNF1A*. *HNF1A* alleles found to be associated with T2D risk consisted mainly of non-coding *HNF1A* variants. Some common and low-frequency variants in coding regions of *HNF1A* are briefly described below.

The recently published GoT2D and T2D-GENES studies, including whole genome sequencing in 2,657 Europeans with and without diabetes and exome sequencing in 12,940 individuals with T2D and non-diabetic controls from five ancestral groups, investigated rare and low-frequency coding variants (Fuchsberger et al. 2016). The sample size was increased by genotyping of further 111,548 individuals and showed that low frequency variants generally did not contribute significantly to the T2D risk. However, variants in *HNF1A* were included in the “monogenic OMIM” group of genes, which as a gene-set was associated with T2D [OR 1.51 (1.25-1.83), $p=2.8 \times 10^{-5}$].

Two distinct association signals at *HNF1A* locus with locus-wide significance [common variant p.I27L, OR 1.10 (1.07-1.13) and low-frequency variant p.A98V, OR 1.21 (1.11-1.31)] had been found by fine-mapping of 39 known T2D susceptibility loci in 27,206 cases and 57,574 controls of European ancestry (Gaulton et al. 2015). Both alleles are coding variants in exon 1 of *HNF1A*.

Another coding *HNF1A* variant c.955G>A, p.G319S was proved to be specific only to the Canadian Oji-Cree indigenous population (only 1 allele reported in GnomAD) in a study including 117 individuals with T2D and 334 non-diabetic subjects (Hegele et al. 1999). Variant p.G319S showed a 4-fold increased risk of T2D in homozygous and 2-fold in heterozygous subjects. Carriers of this variant presented with diabetes earlier than non-carriers (mean age of onset 28.0 ± 5.8 years in homozygotes, 36.6 ± 12.4 years in heterozygotes vs 43.0 ± 11.6 years in non-carriers) and had lower BMI than non-carriers (27.1 vs 30.3 kg/m^2). Variant p.G319S affected splicing of exons 4 and 8 resulting in a truncated mRNA transcript (Harries et al. 2008). Experiments using a HeLa cell line transfected with a plasmid carrying the above *HNF1A* variant had reported transcription activity of the target promoter reduced to 60% of the WT *HNF1A* protein (Triggs-Raine et al. 2002). This functional work and pattern of inheritance supports an association of the allele with multifactorial common disease rather than monogenic diabetes.

A coding *HNF1A* variant c.1522G>A, p.E508K had been associated with an increased prevalence of T2D in a Latino population [OR 5.48 (95% CI 2.83-10.61), $p=4.4 \times 10^{-7}$] (Estrada et al. 2014). This study included 1,794 subjects with T2D and 1,962 non-diabetic

individuals. There was no difference in the clinical characteristics, including age of onset, between diabetic subjects with the p.E508K variant and those without this allele. However, *in vitro* characterisation of the variant using HeLa cells have reported transcription activity reduced to 40% of the WT HNF1A (Estrada, Aukrust et al. 2014).

The association of some *HNF1A* variants with T2D explains a phenotypic spectrum of *HNF1A* alleles which could cause a form of monogenic diabetes (HNF1A-MODY) contribute to the more common T2D, result in the phenotype between those two types of diabetes or no diabetes at all (Flannick, Beer et al. 2013).

3.1.4 Aims of the study

In this chapter, I aimed to:

1. Systematically assess all rare non-synonymous *HNF1A* variant alleles found by sequencing *HNF1A* from 1058 unselected subjects with young adult-onset non-autoimmune diabetes from the UK and Croatia;
2. Calculate an estimated prevalence of HNF1A-MODY in Croatia.

3.2 Study subjects and methods

3.2.1 Study subjects

Subjects participating in this study were recruited from the UK and Croatia. The UK individuals (n=578) were recruited by research nurses via the Young Diabetes in Oxford (YDX) study from seven local diabetes centres and multiple primary care centres as described in chapter 2, section 2.1.1. Croatian subjects (n=472) were recruited by

diabetologists through the Croatian National Registry of individuals with diabetes (CroDiab), as described in chapter 2, section 2.1.2. The studies included subjects with diabetes onset below 45 years and above 18 years of age at recruitment, endogenous insulin production expressed by a fasting C-peptide of more than 0.2nmol/L, and negative GAD antibodies according to the locally established cut-off as listed in chapter 2, section 2.10.2. Co-segregation studies resulted in an assessment of 22 relatives and 8 of 9 relatives with diabetes who fulfilled the entry criteria were also included in the study, bringing the total number of recruited individuals to 1058 (581 from the UK and 477 from Croatia). The clinical characteristics of these subjects are given in table 3.2.1.

Subjects from both countries were of similar age at recruitment and at the diagnosis of diabetes, had similar duration of diabetes, as well as levels of triglycerides, fasting plasma glucose (FPG) and C-peptide (Table 3.2.1). Individuals recruited from Croatia had lower median BMI [29.4 (25.7-33.4) vs 31.7 (27.8-36.7) kg/m², p<0.01], lower HbA1c [(7.3 (6.4-8.3) vs 7.8 (6.7-9.1) %, p<0.01)] and higher total and HDL cholesterol [4.99 (±1.22) vs 4.42 (±1.03) mmol/L and 1.2 (0.9-1.3) vs 1.1 (1.0-1.4) mmol/L, p<0.01] than participants from the UK. Figure 3.2.1 shows the distribution of age at recruitment, age at diagnosis of diabetes and BMI in subjects from both countries.

1058 participants with C-peptide positive, GAD antibody negative diabetes had all 10 exons of *HNF1A* sequenced. The methods were described in chapter 2 and are summarised below.

Participant group depending on country of recruitment	Participants from the UK	Participants from Croatia	P value
Number of subjects	581	477	
Gender (number of men/%)	350 (61.4%)	268 (56.5%)	0.06
Age at recruitment [mean (SD) years]	48.1 (12.5)	47.4 (9.8)	0.37
Age at diagnosis [mean (SD) years]	35.7 (7.5)	35.2 (6.3)	0.31
BMI [median (IQR) kg/m ²]	31.7 (27.8-36.7)	29.4 (25.7-33.4)	3.9 × 10⁻⁹
FPG [median (IQR) mmol/L]	8.2 (6.5-10.5)	8.0 (6.5-9.9)	0.41
HbA1c [median (IQR) %]	7.8 (6.7-9.1)	7.3 (6.4-8.3)	1.5 × 10⁻⁸
Duration of DM [median (IQR) years]	10.0 (4.0-18.0)	11.0 (6.0-16.5)	0.15
C-peptide [median (IQR) nmol/L]	0.67 (0.43-0.99)	0.72 (0.45-0.98)	0.57
Total cholesterol [mean (SD) mmol/L]	4.42 (1.03)	4.99 (1.22)	4.8 × 10⁻¹⁵
HDL [median (IQR) mmol/L]	1.10 (0.90-1.30)	1.20 (1.00-1.40)	6.9 × 10⁻⁷
TG [median (IQR) mmol/L]	1.58 (1.10-2.30)	1.60 (1.14-2.39)	0.38
Treatment			
Insulin +/- OHA [number of case (%)]	271 (46.1)	168 (38.8)	0.02
OHA [number of case (%)]	269 (46.3)	243 (50.9)	0.22
diet [number of case (%)]	41 (7.0)	66 (13.9)	1.0 × 10⁻⁴

Table 3.2.1 Participant characteristics in 2 groups depending on country of recruitment (total n=1058); OHA, oral hypoglycaemic agents and GLP-1 analogues. Significant p value <0.05, Mann-Whitney test to compare variables in both cohorts.

3.2.2 DNA extraction and quantification

Genomic DNA for subjects was extracted from whole blood by Dr Carla Burrows in the Diabetes Research Laboratories in the Oxford Centre for Diabetes, Endocrinology and Metabolism (OCDEM), University of Oxford, UK and by Ms Tamara Pavic in the Zagreb Biochemistry and Molecular Biology Laboratory, University of Zagreb, Croatia (Croatian samples). It was performed on the automated Maxwell 16 system (Promega, Madison, WI, USA) utilising a Blood DNA purification kit (Promega, Madison, WI, USA) in the UK

laboratory as described in chapter 2, section 2.3. The DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany) was used in the Croatian laboratory.

I quantified DNA from all UK and most of the Croatian subjects on a NanoDrop ND-8000 spectrophotometer (ThermoScientific, Wilmington, DE, USA) described in chapter 2, section 2.3.

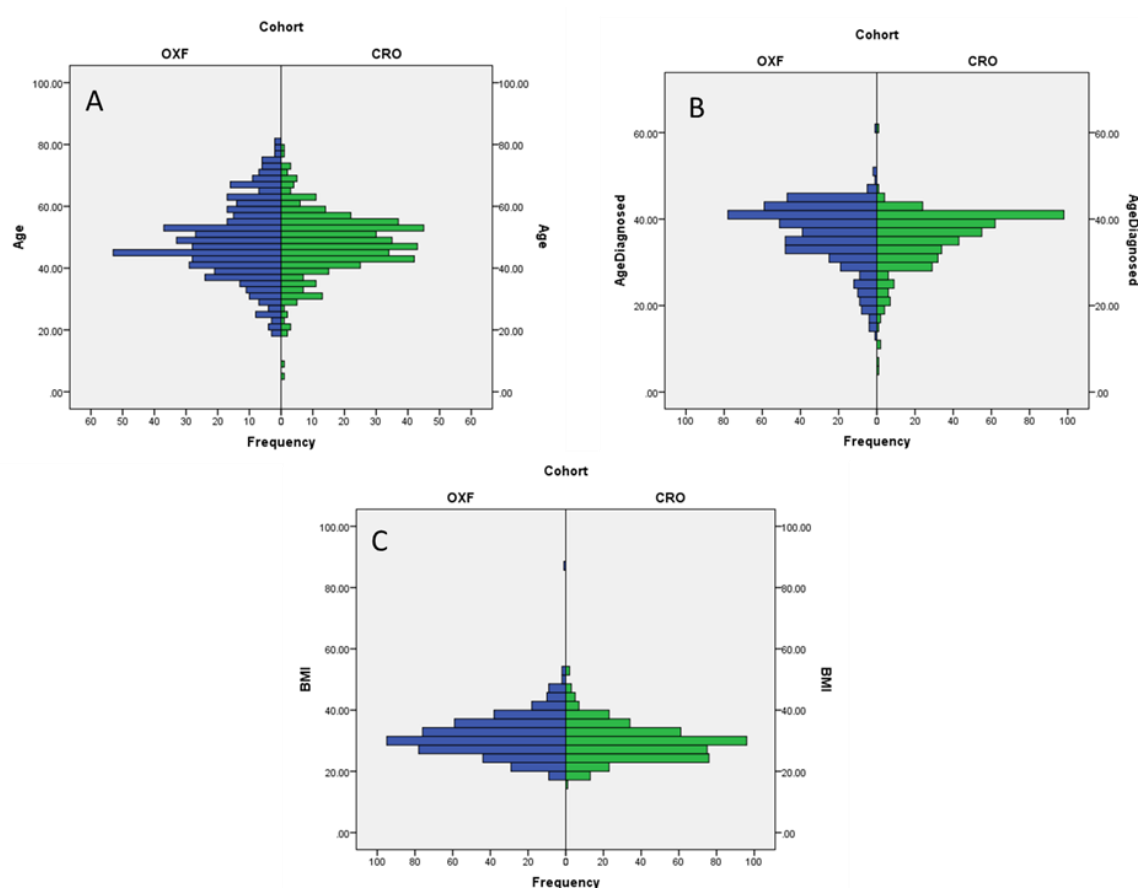


Figure 3.2.1 Distribution of age at recruitment (A), age at diabetes diagnosis (B) and BMI (C) in subjects with the UK (blue) and those from Croatia (green).

3.2.3 DNA amplification and gel electrophoresis

I amplified the genomic DNA from 1058 individuals using a polymerase chain reaction with forward and reverse primers for all 10 *HNF1A* exons. I performed this work, gel electrophoresis, the post-PCR product clean-up and DNA sequencing reactions in the Diabetes Research Laboratories in OCDEM. All methods are described in section 2.4. Primer sequences are listed in Table 3.2.2 and were derived from those previously published (Ellard, Bellanne-Chantelot et al. 2008). PCR reagents and thermocycling conditions are given in chapter 2, Table 2.4.1.

I performed gel electrophoresis for all *HNF1A* amplicons using 2% agarose gels with ethidium bromide to confirm successful amplification of *HNF1A* exons as described in chapter 2, section 2.5. PCR amplification was repeated for all failed samples.

I used Exonuclease I and TSAP for post-PCR clean-up as reported in chapter 2, section 2.6.1.

<i>HNF1A</i> exon	Forward (F) primer all with 5'-TGTAACACGACGGCCAGT M13 tag	Reverse (R) primer all with 5'-CAGGAAACAGCTATGACC M13 tag
1	GCAGTGGGTGCAAGGAGT	GCCCCTCTAGGCTCTCCTG
2	CAAGTCTCTGTCCTCATGACC	TCCCACTGACTTCCTTTCC
3	GGGCAAGGTCAGGGGAATGGA	CAGCCCAGACCAAACCAGCAC
4	CAGATCTGCCAGCCTCAAACC	ACCTTCCCACGTGTCCCTTG
5	TGGAGTTTGAAGTGCTGAGG	GCCAAGGAAAGATGAGGTTG
6	TGTAAGGAAAACCAACCTCA	AGTGGCTCTTCCAGCTCCT
7	GCTCTGGGAAGGAGAGGTG	CATCCATTGACAGCCAACC
8	AGGCTCCCCTGCATCAAG	GGGTGAGGGGCTCTGTCA
9	ACCAAGCAGGTAAGGTCCAG	GTGACGGACAGCAACAGAAG
10	GTACCCCTAGGGACAGGCAGG	ACCCCCAAGCAGGCAGTACA

Table 3.2.2 Forward and reverse primers for all 10 *HNF1A* exons.

3.2.4 Sanger sequencing and annotation of the variants

I used the bidirectional chain-terminating (Sanger) sequencing technique as opposed to next-generation sequencing techniques, because I was interested in only one gene, *HNF1A*. Sanger method was already established in the OCDEM genetic laboratory. I performed the sequencing reaction using the Big Dye Terminator Sequencing mix v1.1 (Applied Biosystems, Birchwood, UK) and M13 sequencing primers which were compatible with M13 tag on the *HNF1A* exons' primers used previously for DNA amplification. The details are described in chapter 2, section 2.6.

I used 96-well Performa DTR Ultra plates (Edge BioSystems, Gaithersburg, USA) for post-sequencing clean-up. The final sequencing read was obtained on an ABI 3730 DNA analyser (Applied Biosystems, Birchwood, UK) by the sequencing facility team of the Department of Zoology, University of Oxford.

I viewed all Sanger sequencing results with Mutation Surveyor DNA Variant Analysis software v5.0.1 (Soft Genetics, Cambridge, UK) and used the human *HNF1A* sequence as a reference file (ID 6927, chromosome 12, NC_000012.11; NM_000545.5; 121415861-121440315) from the NIH GenBank 105 annotation release (GRCh37.p13, GCF_000001405.25). I repeated sequencing for all samples with a sequencing quality score below 30 (range 1-100, reference files usually 70-90) and visually reviewed all electropherograms with quality score 30-40. Sequencing of all rare variants found in the study was repeated from a re-extracted DNA sample for confirmation of the finding.

Twenty six subjects were excluded due to failed sequencing quality of any of 10 *HNF1A* exons leaving 1032 samples for the analysis of sequencing.

3.2.5 Validation of the sequencing results by genotyping

I have genotyped one common (rs1169288, p.I27L) and one low-frequency (rs1800574, p.A98V) *HNF1A* variant from exon 1 to validate the Sanger sequencing results from the first 380 samples (4 96-well plates, 4 samples replaced with negative controls). I chose those two variants as they had been previously reported as associated with T2D and to compare their allele frequency in my study group with young adult onset diabetes and their frequency in GnomAD. I performed PCR using PCR Master Mix (TaqMan, Applied Biosystems, Foster City, CA, USA) and TaqMan SNP genotyping Assay for the examined SNP (Applied Biosystems, Foster City, CA, USA) as given in chapter 2, section 2.7 and Table 2.7.1. The TaqMan ABIPrism 7900HT Sequence Detection System (Applied Biosystems, Foster City, CA, USA) with SDS v2.3 software was utilised for the allelic discrimination end-read.

3.2.5.1 Genotyping of one common and one low-frequency coding HNF1A variant

As a quality control measure, I performed genotyping of one common (c.79A>C, p.I27L, rs1168288) and one low-frequency (c.293C>T, p.A98V, rs1800574) variant from exon 1 of *HNF1A* on 380 samples. The Sanger sequencing failure rate was 4% and all were related to failed PCR of those samples, while the genotyping failure rate was similar in p.A98V (6%) but higher in p.I27L (18%, Figure 3.2.2, Table 3.2.3). I have compared all samples with the assigned genotype by both methods and they had 100% agreement for both variant

Variant allele	Allele count by genotyping	Allele count by Sanger seq.	Discrepancy [%]
c.79A>C, p.I27L			
CC	49 (15.7%)	51 (13.9%)	
AC	126 (40.4%)	148 (40.5%)	
AA	137 (43.9%)	167 (45.6%)	
Total	312	366	
Call rate	82%	96%	0.02%
Failed samples	18%	4% (all failed PCR)	
c.293C>T, p.A98V			
TT	0	0	
CT	30 (8.4%)	32 (8.7%)	
CC	328 (91.6%)	334 (91.3%)	
Total	358	366	
Call rate	94%	96%	0%
Failed samples	6%	4% (all failed PCR)	

Table 3.2.3 Comparison of the call rate and discrepancy rate in genotyping versus Sanger sequencing for 1 common and 1 low-frequency variant (c.79A>C, p.I27L, rs1169288 and c.293C>T, p.A98V, rs1800574).

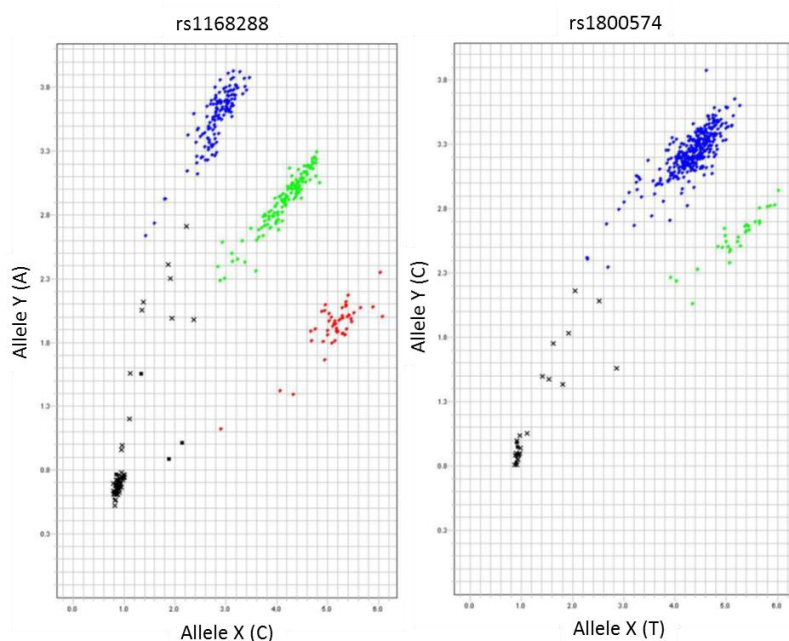


Figure 3.2.2 Allelic discrimination plot from TaqMan genotyping for a common allele: c.79A>C, rs1168288 (left) and low-frequency allele (c.293C>T, rs1800574 (right), allele X (red), both alleles (green), allele Y (blue), undetermined call (black), negative controls (black squares in the left bottom corner of each plot), nucleotide called as X or Y for each allele given in brackets on axes.

alleles with the exception of one sample in p.I27L (0.02% discrepancy, Table 3.2.3). This confirms a correct annotation of the alleles by Mutation Surveyor or myself if quality of sequencing required visual review.

3.2.6 An assessment of the variants using bioinformatic prediction tools

I have used four publically available bioinformatic tools: SIFT (Kumar, Henikoff et al. 2009), Polyphen2 (Adzhubei, Schmidt et al. 2010), Provean (Choi, Sims et al. 2012) and Mutation Taster 2 (Schwarz, Cooper et al. 2014) for *in silico* prediction of the *HNF1A* alleles functional effect (described in chapter 2, section 2.8.2). I chose those four tools and used them consistently throughout my research, being aware that multiple new bioinformatic tools had been developed (Alamut, Condor or CADD). However they all took into account data from the well-established and widely used SIFT and Polyphen 2. I used Mutation Taster 2 first due to its detailed descriptive output including: the position of the mutation in the chromosome; previous reports of the variant; presence of the variant allele in sequencing databases such as 1000 Genomes and ExAC; conservation data comparing amino acid (AA) sequences in humans versus 10 other species; and lists of protein features potentially affected by AA change. However, because this tool was most discordant with predictions of allele effects provided by others and generally predicted alleles to be disease-causing, even if cumulative evidence suggested otherwise. I used SIFT, PolyPhen 2 and Provean for the final assignment of allelic functional effects based on agreement of at least 2/3 of these methods.

3.2.7 Calculating the prevalence of HNF1A-MODY for the Croatian population

First, the proportion of the subjects with likely damaging *HNF1A* alleles (probable HNF1A-MODY) in the total number of participants in the study was calculated and extrapolated to the population of subjects with T2D in Croatia diagnosed before 45 years of age as reported by the Croatian Institute of Public Health in 2015 (http://www.idb.hr/web_english/crodiab.htm; accessed October 12, 2016). Subsequently, the number of individuals with HNF1A-MODY in a population of subjects with T2D diagnosed before 45 years of age was used to calculate the number of individuals with HNF1A-MODY per one million of population in Croatia in 2015 (www.countrymeters.info/en/Croatia/), and thereby a prevalence of HNF1A-MODY in Croatia. The estimate of the population in Croatia was based on the last census from 2011. The 95% confidence interval (CI) was calculated using the following equation:

$$p \pm CI = p \pm 1.96 \times \sqrt{\frac{p \times (1 - p)}{n}}$$

where p is the mean proportion of subjects with disease in the given population (n) (<https://onlinecourses.science.psu.edu/stat100/node/56>).

3.3 Results

3.3.1 *HNF1A* variants found in the study

Bidirectional sequencing of all 10 exons of *HNF1A* from 1032 individuals with young adult onset non-autoimmune diabetes identified 26 rare (MAF<1% in ExAC) non-synonymous

HNF1A variants (25 coding and 1 non-coding promoter variant, Figure 3.3.1), 1 low-frequency (MAF 1-5%) and 3 common (MAF>5%) non-synonymous coding *HNF1A* variants. Additionally, 8 synonymous coding *HNF1A* variants were found.

The design of the primers for *HNF1A* exons allowed the identification of 17 intronic (non-coding) variants in the regions adjacent to exons of *HNF1A* (± 85 bp depending on the exon-specific primer). All variants identified in this study are listed in Tables 3.3.1, 3.3.2 and 3.3.3 and described below.

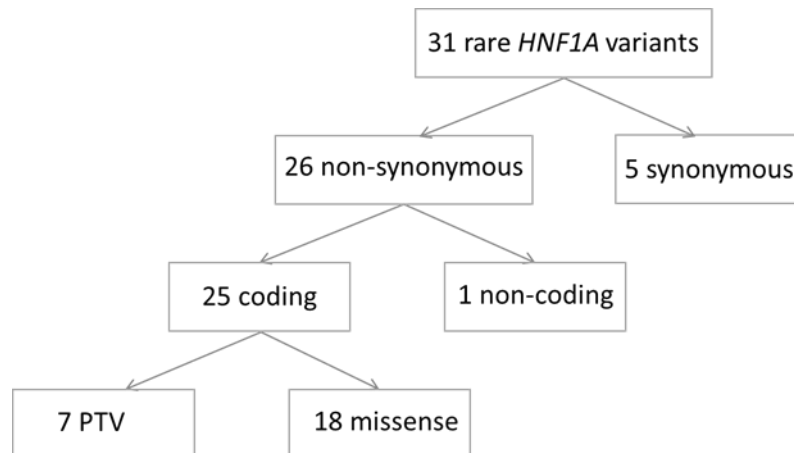


Figure 3.3.1 Rare *HNF1A* variants found in this study; PTV, protein truncating variants.

3.3.1.1 Coding *HNF1A* variants

Twenty five rare non-synonymous coding *HNF1A* variants were found in 29 probands and 8 relatives (Table 3.3.1). Twenty two of the variants were located in exons 1-6, thus affecting all three main isoforms of the protein. Three variants were in exons 8-10, hence only affecting isoform A which is predominant in the liver and kidneys but not in the pancreatic islets. Exons 4 and 6 harboured most of the variants (6 in 14 individuals and 5

in 7 individuals respectively), while there were no variants reported in exon 7. Eleven variants were in the dimerization or DNA binding domains and 14 in the transactivation domain. The most frequently affected codons were P291 and P379 (in 5 individuals each), both corresponding to the poly-C regions of the exon and reported as most prone to genetic variation in the literature (Colclough, Bellanne-Chantelot et al. 2013). The listed in Table 3.3.1 rare non-synonymous coding *HNF1A* variants included 7 protein truncating variants (PTV) and 18 missense variants.

Of 8 common or low frequency coding *HNF1A* variants, 4 were non-synonymous and 4 were synonymous. The remaining 5 synonymous coding *HNF1A* variants were rare (Table 3.3.2).

DNA variant	Chromosomal position [#]	Protein variant	Exon	Domain	Isoform	Variant type	MAF in this study	MAF in GnomAD
Non-synonymous rare variants								
c.1_326del	121416572-121416897	del exon 1	1	Dimerisation	1, 2, 3	exon del	n/a	0
c.8C>G	121416579	S3C	1	Dimerisation	1, 2, 3	missense	0.0005	0
c.139G>C	121416710	G47R	1	Dimerisation	1, 2, 3	missense	0.0005	0.00001
c.142G>A	121416713	E48K	1	Dimerisation	1, 2, 3	missense	0.0005	0.00009
c.404delA	121426713	D135fs	2	DNA binding	1, 2, 3	frameshift	0.0005	0
c.451G>A	121426760	G151S	2	DNA binding	1, 2, 3	missense	0.0005	0
c.586A>G	121431382	T196A	3	DNA binding	1, 2, 3	missense	0.0005	0.00027
c.666G>T	121431462	K222N	3	DNA binding	1, 2, 3	missense	0.0005	0
c.685C>T	121431481	R229*	3	DNA binding	1, 2, 3	nonsense	0.001	0.000008
c.751G>A	121432004	A251T	4	DNA binding	1, 2, 3	missense	0.0014	0
c.779C>T	121432032	T260M	4	DNA binding	1, 2, 3	missense	0.0014	0
c.862G>T	121432115	G288W	4	Transactivation	1, 2, 3	missense	0.001	0.00007
c.871C>A	121432124	P291T	4	Transactivation	1, 2, 3	missense	0.0014	0.000008
c.872delC	121432125	P291fs	4	Transactivation	1, 2, 3	frameshift	0.001	0
c.872dupC	121432125	G292fs	4	Transactivation	1, 2, 3	frameshift	0.0014	0
c.1015G>A	121434124	G339S	5	Transactivation	1, 2, 3	missense	0.0005	0.0002
c.1047C>A	121434156	H349Q	5	Transactivation	1, 2, 3	missense	0.0005	0.00006
c.1129delC	121434365	L377fs	6	Transactivation	1, 2, 3	frameshift	0.0005	0
c.1136_1137delCT	121434372-121434373	P379fs	6	Transactivation	1, 2, 3	frameshift	0.001	0
c.1136C>A	121434372	P379H	6	Transactivation	1, 2, 3	missense	0.0005	0.00005
c.1136C>G	121434372	P379R	6	Transactivation	1, 2, 3	missense	0.001	0
c.1165T>G	121434401	L389V	6	Transactivation	1, 2, 3	missense	0.0005	0.0006
c.1544C>T	121437113	T515M	8	Transactivation	1	missense	0.0005	0.00002
c.1748G>A	121437382	R583Q	9	Transactivation	1	missense	0.0005	0.0005
c.1816G>A	121438915	G606S	10	Transactivation	1	missense	0.0005	0.00005

Table 3.3.1 Rare (MAF<0.01 in ExAC) non-synonymous coding HNF1A alleles found in my study.

[#]Chromosomal position from human genome assembly GRCh37 Ensembl 63, canonical transcript ENST00000257555; ^{\$} allele frequency in non-Finish Europeans in GnomAD (expressed as a proportion).

DNA variant	Chromosomal position#	Protein variant	Exon	Protein domain	Variant type	Allele frequency in this study	Allele frequency in GnomAD [§]
Non-synonymous common & low frequency variants							
c.79A>C	121416650	I27L	1	Dimerisation	missense	0.34 (C)	0.34 (C)
c.293C>T	121416864	A98V	1	Dimerisation	missense	0.05 (T)	0.03 (T)
c.1460G>A	121435427	S487N	7	Transactivation	missense	0.32 (A)	0.31 (A)
c.1720A>G	121437382	S574G	9	Transactivation	missense	0.99 (G)	0.99 (G)
Synonymous rare variants							
c.276C>T	121416847	L92L	1	Dimerisation	synonymous	0.0005 (T)	0.001 (T)
c.693G>A	121431489	T231T	3	DNA binding	synonymous	0.0002 (A)	0.0002 (A)
c.1386C>T	121435353	V462V	7	Transactivation	synonymous	0.0005 (T)	0.0003 (T)
c.1611G>A	121437180	T537T	8	Transactivation	synonymous	0.0005 (A)	0.000007 (A)
c.1875G>A	121438974	Q625Q	10	Transactivation	synonymous	0.0005 (A)	0
Synonymous common variants							
c.52C>G	121416622	L17L	1	Dimerisation	synonymous	0.46 (G)	0.47 (G)
c.864G>C	121432117	G288G	4	DNA binding	synonymous	0.27 (C)	0.30 (C)
c.1375C>T	121435342	L459L	7	Transactivation	synonymous	0.32 (T)	0.31 (T)
c.1545G>A	121437114	T515T	8	Transactivation	synonymous	0.16 (A)	0.19 (A)

Table 3.3.2 Common, low-frequency non-synonymous and rare synonymous coding *HNF1A* variants detected by sequencing in my study. [#]Chromosomal position from human genome assembly GRCh37 Ensembl 63, canonical transcript ENST00000257555; [§] allele frequency in non-Finnish Europeans in GnomAD (given as a proportion).

3.3.1.2 Non-coding *HNF1A* variants

There were 17 non-coding *HNF1A* variants, including 13 common, 3 low frequency and one rare variant, which was identified in the *HNF1A* promoter region in one proband (Table 3.3.3). This promoter missense variant is discussed together with all rare non-synonymous coding *HNF1A* variants in all sections below. The minor allele frequency for all common variants was comparable to the MAF in non-Finnish Europeans in GnomAD and could be used as a measure of quality control for my Sanger sequencing. There were two intronic variants reported in GnomAD as rare or low-frequency but in my study had MAF consistent with a common variant (c.526+66C>G, MAF 18% and c.1768+44C>T, MAF 11%). The MAF of the first variant was 0.007% and the second variant was of low

frequency (5%) in non-Finnish European subgroup of GnomAD. The difference in MAF could be due to an increased allele frequency in subjects with T2D.

DNA variant	Chromosomal position	Intron	Variant type	Allele frequency in this study	Allele frequency in GnomAD ^{\$}
c.-4A>G	121416568	5'UTR	Promoter	0.0005 (G)	0.0007 (G)
c.327-158C>T	121426478	1	intronic	0.36 (T)	0.31 (T)
c.327-42G>A	121426594	1	intronic	0.34 (A)	0.33 (A)
c.526+66C>G	121426888	2	intronic	0.18 (G)	0.00007 (G)
c.527-51T>A	121431272	2	intronic	0.23 (A)	0.23 (A)
c.527-23C>T	121431300	2	intronic	0.32 (T)	0.31 (T)
c.1108-42G>T	121434302	5	intronic	0.02 (T)	0.02 (T)
c.1108-27G>T	121434317	5	intronic	0.01 (T)	0.005 (T)
c.1501+7G>A	121435475	7	intronic	0.37 (A)	0.36 (A)
c.1623+29T>C	121437221	8	intronic	0.79 (C)	0.81 (C)
c.1623+35G>A	121437227	8	intronic	0.21 (A)	0
c.1624-69C>T	121437217	8	intronic	0.03 (T)	0
c.1624-67C>T	121437219	8	intronic	0.06 (T)	0
c.1768+44C>T	121437474	9	intronic	0.11 (T)	0.05 (T)
c.1768+85T>G	121437525	9	intronic	0.21 (G)	0
c.1769-24T>C	121438844	9	intronic	0.37 (C)	0.37 (C)
c.1896+47A>G	121439042	10	intronic	0.05 (G)	0

Table 3.3.3 Non-coding *HNF1A* variants detected by sequencing in my study. [#]Chromosomal position from human genome assembly GRCh37 Ensembl 63, canonical transcript ENST00000257555; ^{\$} allele frequency in non-Finish Europeans in GnomAD (expressed as a proportion).

3.3.2 Bioinformatic prediction for rare non-synonymous *HNF1A* variants

I used three bioinformatic tools (SIFT, Polyphen 2 and Provean) for *in silico* prediction of the allele functional effect as described earlier. All protein truncating variants were in exons 1-6 (present in all 3 main isoforms of the *HNF1A* protein) and assumed to be disease-causing due to the likelihood of nonsense-mediated decay leading to haploinsufficiency (Rivas et al. 2015). Provean provided a prediction for PTV and predicted all to be damaging. However, there were no PTVs in exons 7-10 in this study.

At least 2 of the 3 bioinformatic tools predicted that 7/18 rare missense coding variants are likely to be damaging to protein function, 11/18 likely to be benign. Human Splicing Finder and Mutation Taster 2 also predicted that a non-coding missense variant in the promoter, c.-4A>G would not affect the splicing of exon 1 of *HNF1A* from the promoter (Table 3.3.4). Both tools predicted that the novel variant c.8C>G, p.S3C is likely to influence splicing of the exon 1 from promoter by gaining a new donor site as well as causing a coding change.

3.3.3 Review of previous reports in literature

Fourteen of twenty six rare non-synonymous *HNF1A* variants had been previously reported in the literature as MODY-causing, 9 had been reported as variants of unknown significance (VUS) or benign (Colclough, Bellanne-Chantelot et al. 2013) and 3 were novel (p.S3C, p.G151S, p.K222N) (Table 3.3.5).

Coding DNA variant	Protein variant	Variant type	Reported as causing MODY	Bioinformatics prediction		
				SIFT	Polyphen 2	Provean
c.-4A>G	n/a	missense	no	n/a		
c.1-326del	del exon 1	exon del	yes	protein truncating		
c.8C>G	S3C	missense	novel	damaging	possibly damaging	deleterious
c.139G>C	G47R	missense	yes	tolerated	benign	neutral
c.142G>A	E48K	missense	yes	tolerated	benign	neutral
c.404delA	D135fs	frameshift	yes	protein truncating		
c.451G>A	G151S	missense	novel	damaging	probably damaging	deleterious
c.586A>G	T196A	missense	no	tolerated	benign	neutral
c.666G>T	K222N	missense	novel	damaging	probably damaging	deleterious
c.685C>T	R229*	nonsense	yes	protein truncating		
c.751G>A	A251T	missense	no	tolerated	probably damaging	neutral
c.779C>T	T260M	missense	yes	damaging	probably damaging	deleterious
c.862G>T	G288W	missense	no	tolerated	probably damaging	neutral
c.871C>A	P291T	missense	yes	tolerated	benign	neutral
c.872delC	P291fs	frameshift	yes	protein truncating		
c.872dupC	G292fs	frameshift	yes	protein truncating		
c.1015G>A	G339S	missense	no	tolerated	benign	neutral
c.1047C>A	H349Q	missense	no	tolerated	benign	neutral
c.1129delC	L377fs	frameshift	yes	protein truncating		
c.1136_1137delCT	P379fs	frameshift	yes	protein truncating		
c.1136C>A	P379H	missense	yes	damaging	probably damaging	deleterious
c.1136C>G	P379R	missense	yes	damaging	probably damaging	deleterious
c.1165T>G	L389V	missense	no	tolerated	possibly damaging	neutral
c.1544C>T	T515M	missense	no	damaging	probably damaging	neutral
c.1748G>A	R583Q	missense	no	tolerated	benign	neutral
c.1816G>A	G606S	missense	yes	tolerated	benign	neutral

Table 3.3.4 Bioinformatic prediction on functional effect of rare non-synonymous *HNF1A* variants found in my study. Alleles predicted as damaging protein function highlighted in pink (7 PTVs & 7 missense variants).

3.3.4 Co-segregation of the rare non-synonymous *HNF1A* alleles and diabetes

Cascade screening of the 22 available family members from 6 probands (variants p.R229*, p.A251T, p.T260M, p.G288W, p.P291T and p.P379R) confirmed the same *HNF1A* variant allele as in the proband in 8 individuals. Co-segregation of the *HNF1A* allele with diabetes in the family has been reported in literature or in this study in case of 15 alleles (Table

3.3.5). An inheritance pattern of diabetes consistent with autosomal dominant disease was seen in families of the subjects with three additional alleles, but relatives were not available for sequencing. Due to a small number of relatives available for sequencing (1-4 per family), there was no power to exclude co-segregation of the *HNF1A* allele with diabetes in any family.

Coding DNA variant	Protein variant	Reported as causing MODY*	Co-segregation of the variant with DM
c.-4A>G	n/a	no	not available
c.1-326del	del exon 1	yes	yes (Ellard 2007)
c.8C>G	S3C	novel	not available
c.139G>C	G47R	yes	yes (Thanabalasingham 2013)
c.142G>A	E48K	yes	yes (Moller 1998)
c.404delA	D135fs	yes	yes (Frayling 2001)
c.451G>A	G151S	novel	not available
c.586A>G	T196A	no	no (Galan 2011)
c.666G>T	K222N	novel	3 generations DM, only proband sequenced
c.685C>T	R229*	yes	yes (Kawasaki 1997)
c.751G>A	A251T	no	yes (Pavic 2016)
c.779C>T	T260M	yes	yes (Lehto 1997), this study
c.862G>T	G288W	no	no (this study)
c.871C>A	P291T	yes	yes (Thanabalasingham 2011), this study
c.872delC	P291fs	yes	yes (Vaxillaire 1997)
c.872dupC	G292fs	yes	yes (Lehto 1997)
c.1015G>A	G339S	no	yes (Mughal, unpublished)
c.1047C>A	H349Q	no	not available
c.1129delC	L377fs	yes	2 generations DM, only proband sequenced
c.1136_1137delCT	P379fs	yes	yes (Kawasaki 2000)
c.1136C>A	P379H	yes	yes (Dominguez-Lopez 2005)
c.1136C>G	P379R	yes	yes (Xu 2005)- 2 gen, this study
c.1165T>G	L389V	no	no co-segregation provided
c.1544C>T	T515M	no	not available
c.1748G>A	R583Q	no	yes (Yamada 1997), 2 gen included
c.1816G>A	G606S	yes	2 generations DM, only proband sequenced

Table 3.3.5 Rare non-synonymous *HNF1A* alleles (25 coding and one non-coding promoter variant) from this study reported in previous literature (PubMed search) including evidence of co-segregation. Alleles previously reported as causing MODY marked in pink; *reported as MODY-causing in the available literature; gen, generation.

3.3.5 Clinical features of subjects with rare non-synonymous *HNF1A* variants

The clinical characteristics of an individual can guide a clinician in choosing those with a high probability of MODY for genetic testing (Chapter 1). Thus, young age of diabetes onset (≤ 30 years of age), normal BMI, sensitivity to sulphonylurea or glinide would suggest *HNF1A*-MODY and can contribute to the assessment of the allele's functional effect.

In my study, 17 of 38 subjects (including 15 of 30 probands) with rare non-synonymous *HNF1A* variants have been diagnosed with diabetes ≤ 30 years of age, 12 individuals (10 probands) had BMI in the normal range ($< 25 \text{ kg/m}^2$) and 21 subjects (all probands) were treated with sulphonylurea (SU) or glinide (mono- or combined therapy) with available evidence for good glycaemic control on a small dose of SU in 10 subjects (mean HbA1c 6.8%) (Table 3.3.6).

The Exeter MODY probability calculator provides a pre-test probability of MODY based on clinical features (age at diagnosis and current age, gender, current treatment, BMI, HbA1c and parental history of diabetes) to guide genetic testing. However, it cannot be used for subjects diagnosed with diabetes above 35 years of age as its performance worsened above that age (Shields, McDonald et al. 2012). The MODY probability cut-off 20% had sensitivity of 97% and specificity of 82%.

This calculator gave $> 20\%$ probability of MODY to 13 of 38 individuals with rare *HNF1A* alleles in this study including 10 subjects with likely damaging *HNF1A* alleles. The calculator was unable to assess 11 subjects due to age at diagnosis > 35 years (including

one with PTVs) and three due to missing age of diabetes onset. The calculator gave <15% probability of MODY to four other subjects with *HNF1A* protein truncating variants.

Study_ID	Coding DNA variant	Protein change	Clinical features					
			Age of onset [years]	BMI [kg/m ²]	Generations with DM	HbA1c	On SU/glinide	Probability of MODY (Exeter calculator) [%]
2047_003	c.-4A>G	c.-4A>G	37	28.4	1	9.7	yes+metf	too old
0555_003	c.1-326del	del exon 1	20	34.2	2	10.1	yes+ins	>21.0
MCRO_496	c.8C>G	S3C	37	24.2	1	8.8	yes+ins	too old
0647_003	c.139G>C	G47R	37	28.8	2	7.6	yes+metf	too old
0587_003	c.142G>A	E48K	42	39.7	2	11.9	no	too old
2175_003	c.404delA	D135fs	28	19.3	3	6.7	yes	>75.5
MCRO_042	c.451G>A	G151S	35	26.3	2	6.3	yes+metf+ins	>15
2114_003	c.586A>G	T196A	42	38.9	2	10.1	no	too old
2566_003	c.666G>T	K222N	18	32	3	7.3	yes	>75.5
MCRO_294	c.685C>T	R229*	25	19.05	3	7.3	yes	>75.5
MCRO_466 [#]	c.685C>T	R229*	36	n/a	3	6.7	no	too old
MCRO_245	c.751G>A	A251T	39	27.1	2	6.8	no	too old
MCRO_343 [#]	c.751G>A	A251T	50	23.8	2	5.3	no	too old
MCRO_345 [#]	c.751G>A	A251T	n/a	23.9	2	5.8	no	n/a
2162_003	c.779C>T	T260M	25	30.5	3	7.8	yes	>45.5
2192_003 [#]	c.779C>T	T260M	18	29.6	3	7.1	no	>62.4
MCRO_083	c.862G>T	G288W	39	38.4	2	5.5	no	too old
MCRO_337 [#]	c.862G>T	G288W	n/a	n/a	2	6	n/a	n/a
0522_003 [#]	c.871C>A	P291T	15	32.5	2	9.5	metf+ins	>45.5
0561_003	c.871C>A	P291T	27	38.8	2	8.6	ins	>4.6
0699_003 [#]	c.871C>A	P291T	33	37	2	9	metf+ins	>4.6
0617_003	c.872delC	P291fs	30	32	2	7.7	yes	>4.6
0785_003	c.872delC	P291fs	18	25.8	3	5.7	yes	>75
0962_003	c.872dupC	G292fs	12	24.9	2	11.3	yes+ins	>2.6
0967_003	c.872dupC	G292fs	17	21	1	6.1	yes	>75.5
MCRO_379	c.872dupC	G292fs	23	24	n/a	6.2	yes	>2.6
2024_003	c.1015G>A	G339S	26	29.7	3	8.3	yes	>32.9
2510_003	c.1047C>A	H349Q	33	n/a	1	9.4	no	>4.6
0505_003	c.1129delC	L377fs	27	26.8	2	6.7	yes	>45.5
0957_003	c.1136_1137delCT	P379fs	35	23.5	3	6.6	yes	>32.9
MCRO_380	c.1136_1137delCT	P379fs	16	22.3	n/a	6.5	no	>6.4
2196_003	c.1136C>A	P379H	41	30.7	3	7.3	yes	too old
MCRO_055	c.1136C>G	P379R	28	19.9	3	7.6	yes+ins	>49.4
MCRO_385 [#]	c.1136C>G	P379R	n/a	29.7	2	6.8	no	n/a
2550_003	c.1165T>G	L389V	37	31.5	2	9.2	yes+metf+ins	too old
MCRO_072	c.1544C>T	T515M	33	27.9	2	7	no	>15.1
2035_003	c.1748G>A	R583Q	30	38.5	2	9	yes	>4.6
0712_003	c.1816G>A	G606S	32	24	2	6.7	no	>4.6

Table 3.3.6 Clinical characteristics of subjects with rare non-synonymous *HNF1A* alleles from this study (n=38, probands and relatives[#] included). BMI, body mass index; SU, sulphonylurea; n/a, not available; metf, metformin; ins, insulin; features supportive of MODY phenotype shaded in pink, subjects shaded in green had documented sensitivity to the small dose of SU.

3.3.6 Allele frequency of the rare non-synonymous *HNF1A* alleles in large-scale sequencing samples

I reviewed the minor allele frequency (MAF) of all rare non-synonymous *HNF1A* variants found in my study in two large-scale sequencing projects: ExAC (n=60,706 exomes; including data of around 36,000 Europeans) and GnomAD (n=123,136 exomes and 15,496 whole-genomes; including data of around 63,000 non-Finish Europeans) created by an expansion of ExAC. I also searched for the alleles identified in this study in the T2D-GENES project which contributed to the ExAC dataset and separated non-diabetic individuals (n=6,436 exomes) from those with T2D.

Based on the prevalence of *HNF1A*-MODY of ~60 cases/1,000,000 of general population from UK, Norway and Croatia (this study),(Eide, Raeder et al. 2008, Shields, Hicks et al. 2010), I chose the MAF cut off >0.007% (>14/245,340 alleles; >7 individuals /~123,000 exomes/genomes) to be too common for an allele causing a typical MODY phenotype. Fourteen of 26 rare non-synonymous *HNF1A* alleles found in my study were present in at least 2 exomes in GnomAD (all ethnic groups included). All but 2 variant alleles (2 homozygotes with allele c.-4A>G, one with p.349Q) were only present in heterozygotes. Six of 26 rare *HNF1A* alleles had MAF>0.007% in GnomAD and five of those were also reported in at least four non-diabetic individual from the T2D-GENES study (Table 3.3.6). Only 4 of 26 alleles had MAF>0.007% when non-Finnish individuals in GnomAD were analysed.

There were three variant alleles which were ethnic group specific (p.G339S and p.H349Q reported only in South Asians; p.L389V only in Africans) and absent in European subgroup of GnomAD. All three probands carrying those alleles in my study were also of South Asian or African origin. Variant c.-4A>G and p.G606S were also more common in South Asian than European individuals (Table 3.3.7).

Most MODY-causing alleles (n=8) and three novel variants were not present in any of the chosen large-scale sequencing datasets. Variant p.R229* was the only protein truncating variant present in GnomAD (MAF 0.0008% in all GnomAD individuals). Figure 3.3.2 plots number of cases with rare non-synonymous *HNF1A* variants found in this study and reported previously in MODY database (Colclough, Bellanne-Chantelot et al. 2013) against number of cases with the same *HNF1A* variants in GnomAD. It showed that the MODY database was enriched for the alleles causing HNF1A-MODY while the benign alleles were more often reported in GnomAD.

Coding DNA variant	Protein change	MAF in all in ExAC [%]	T2D-GENES Controls	MAF in all in GnomAD [%]	MAF in non-Finnish Europeans in GnomAD [%]
c.-4A>G	c.-4A>G	0.1	44	0.08 (0.61 SA)	0.007
c.1-326del	del exon 1	0	0	0	0
c.8C>G	S3C	0	0	0	0
c.139G>C	G47R	0.001	0	0.001	0.002
c.142G>A	E48K	0.02	1	0.009	0.02
c.404delA	D135fs	0	0	0	0
c.451G>A	G151S	0	0	0	0
c.586A>G	T196A	0.03	4	0.03	0.05
c.666G>T	K222N	0	0	0	0
c.685C>T	R229*	0.002	0	0.0008	0.0009
c.751G>A	A251T	0	0	0	0
c.779C>T	T260M	0	0	0	0
c.862G>T	G288W	0.01	0	0.007	0.01
c.871C>A	P291T	0.002	0	0.0008	0.0009
c.872delC	P291fs	0	0	0	0
c.872dupC	G292fs	0	0	0	0
c.1015G>A	G339S	0.02	4	0.02 (0.05 SA)	0
c.1047C>A	H349Q	0.007	3	0.006 (0.05 SA)	0
c.1129delC	L377fs	0	0	0	0
c.1136_1137delCT	P379fs	0	0	0	0
c.1136C>A	P379H	0.003	1	0.005	0.002
c.1136C>G	P379R	0	0	0	0
c.1165T>G	L389V	0.05	16	0.06 (0.66 Af)	0
c.1544C>T	T515M	0.002	0	0.002	0.003
c.1748G>A	R583Q	0.05	5	0.05	0.09
c.1816G>A	G606S	0.005	0	0.005 (0.03 SA)	0.0009

Table 3.3.7 Presence of the rare non-synonymous HNF1A alleles from this study in large-scale sequencing datasets: ExAC, T2D-GENES and GnomAD with minor allele frequency (MAF). MAF in non-Finnish Europeans in GnomAD given separately; variant alleles more common in South Asians (SA) or Africans (Af) shaded in grey.

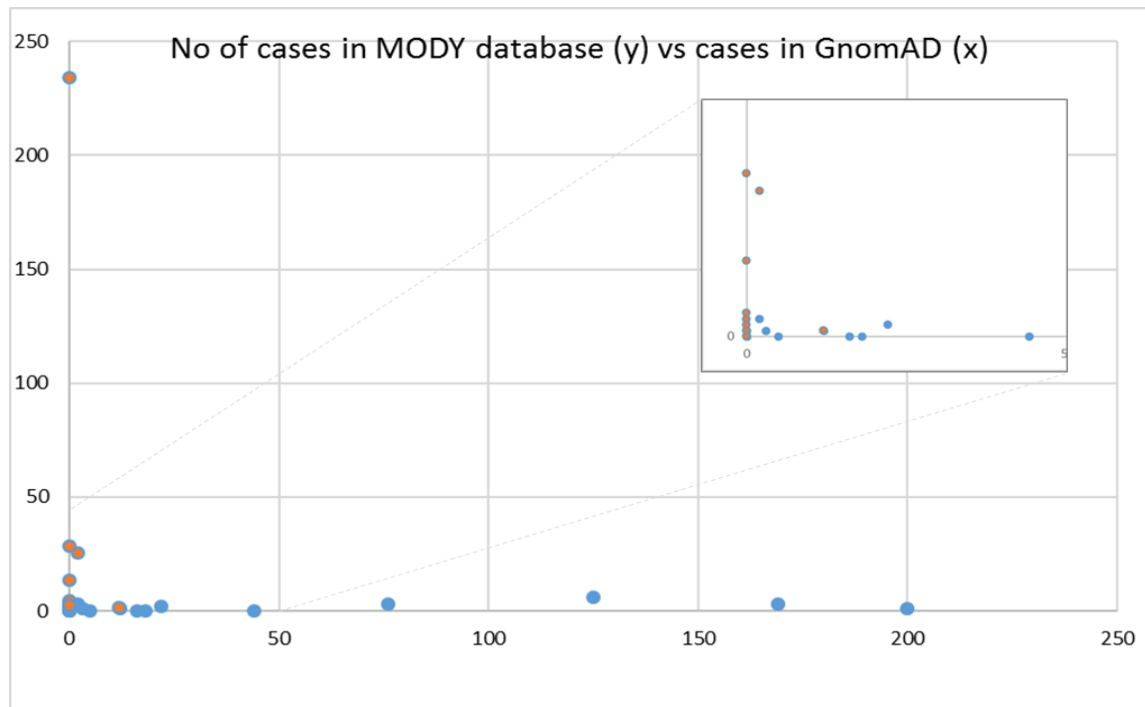


Figure 3.3.2 Number of cases with rare non-synonymous *HNF1A* variants found in this study and in MODY database (y axis) (Colclough, Bellanne-Chantelot et al. 2013) plotted against number of cases with the same *HNF1A* rare variants in GnomAD (x axis), in orange disease-causing variants, in blue benign variants and variants of unknown significance.

Recent recommendations for interpretation of genetic variation include taking into account presence of the allele in the large-scale sequencing projects (MacArthur, Manolio et al. 2014, Quintans, Ordonez-Ugalde et al. 2014). Hence, I have included allele frequency in GnomAD in my overall prediction of the functional effect of *HNF1A* alleles. For that reason, I used an ethnic group specific MAF in GnomAD for the given individual.

3.3.7 Prediction of the phenotypical effect of rare *HNF1A* alleles based on aspects presented

Based on the assessment of all evidence for disease-causality or neutral effect of rare non-synonymous *HNF1A* alleles found in my study, I designed an experimental score which ranges from 0 to 3 and allocates 1 point for each feature of the variant analysed supporting diagnosis of MODY:

- reported as MODY-causing with co-segregation of the variant with diabetes (1),
- predicted by bioinformatic tools as damaging (1),
- MAF in GnomAD (in ethnic specific group) <0.007% and absence of the allele in non-diabetic controls from T2D-GENES subjects (1),

A score of 0-1 was defined as benign, 1.5 likely benign, 2 likely damaging and 2.5-3 damaging variant allele (Table 3.3.8). This prediction score may seem arbitrary, however it does not prioritise one feature over another providing a broader approach as opposed to “a clinically-based expert opinion”. The difference between my scoring system and *in silico* prediction tools (such as Alamut or Condor) is the addition of co-segregation to improve the prediction. An even spread of the variants between benign, through likely benign and likely damaging to damaging reflects the functional spectrum of the genetic variation and highlights the continuum of the allelic effect rather than binary outcome which will be discussed further in chapter 7. The clinical features of subjects in this study were excluded from the score as individuals with the same variant allele may have

different phenotypes depending on the penetrance of the variant and environmental factors.

Using the above score, all PTVs had a score of >2.5 (damaging). For some of the variants, co-segregation data were not available (e.g. variant p.S3C and p.T515M) and the score may have been underestimated.

I concluded that the assessment of genetic variation is complex, which is well-depicted by the Venn diagram below (Figure 3.3.3). The above systematic assessment of rare non-synonymous *HNF1A* alleles found in my study and the provisional functional prediction score depict well the functional spectrum of genetic variation. However, for the purpose of biomarker assessment described in chapter 5, I will reassess the *HNF1A* rare alleles after adding the results of the functional studies discussed in chapter 4.

At current stage of prediction, a total of 10 of 26 alleles (in 14 probands and 3 relatives) are predicted as damaging; 6 of 26 alleles as likely damaging (in 6 probands and 2 relatives), 1 alleles (in 1 proband and 2 relatives) are labelled as likely benign and 9 alleles (in 9 probands and 1 relative) are predicted to be benign (Table 3.3.8).

Coding DNA variant	Protein variant	Reported as causing MODY with co-segregation	<i>In silico</i> prediction	GnomAD MAF<0.007 % or absense in T2D-GENES	Final score	Final prediction
c.-4A>G	n/a	0	0	0	0	benign
c.1-326del	del exon 1	1	1	1	3	damaging
c.8C>G	S3C	novel	1	1	2	likely damaging
c.139G>C	G47R	1	0	1	2	likely damaging
c.142G>A	E48K	1	0	0	1	benign
c.404delA	D135fs	1	1	1	3	damaging
c.451G>A	G151S	novel	1	1	2	likely damaging
c.586A>G	T196A	0	0	0	0	benign
c.666G>T	K222N	novel	1	1	2	likely damaging
c.685C>T	R229*	1	1	1	3	damaging
c.751G>A	A251T	0.5	0	1	1.5	likely benign
c.779C>T	T260M	1	1	1	3	damaging
c.862G>T	G288W	0	0	1	1	benign
c.871C>A	P291T	1	0	1	2	likely damaging
c.872delC	P291fs	1	1	1	3	damaging
c.872dupC	G292fs	1	1	1	3	damaging
c.1015G>A	G339S	0	0	0	0	benign
c.1047C>A	H349Q	0	0	0	0	benign
c.1129delC	L377fs	1	1	1	3	damaging
c.1136_1137delCT	P379fs	1	1	1	3	damaging
c.1136C>A	P379H	1	1	1	3	damaging
c.1136C>G	P379R	1	1	1	3	damaging
c.1165T>G	L389V	0	0	0	0	benign
c.1544C>T	T515M	0	1	1	2	likely damaging
c.1748G>A	R583Q	0	0	0	0	benign
c.1816G>A	G606S	0	0	0	0	benign

Table 3.3.8 Proposed provisional prediction score of the allele functional effect based on the four features discussed above and before results of functional studies considered. Provisional prediction score 0-4 (0-1, benign; 1.5, likely benign; 2, likely damaging; 2.5-3, damaging allele); #MAF in GnomAD in the ethnic specific group for an individual.

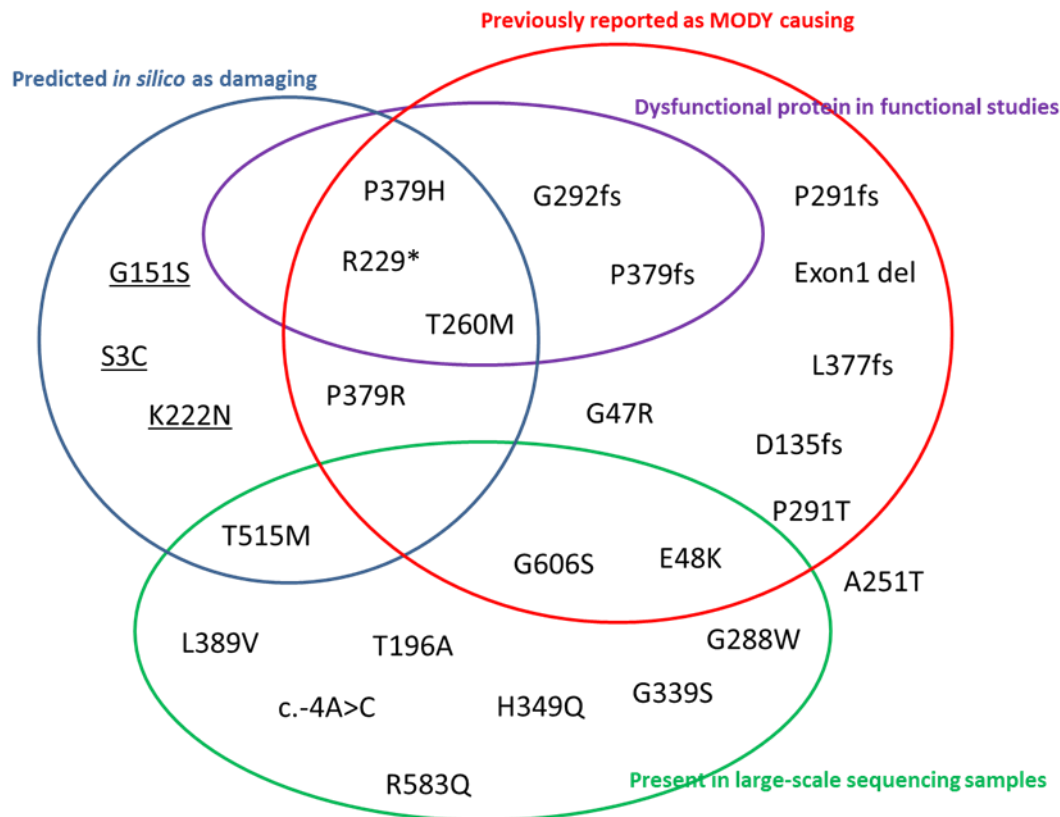


Figure 3.3.3 Venn diagram illustrating the complexity of assigning disease causality of the rare non-synonymous HNF1A variants. Variants predicted by in silico tools as damaging (blue circle), previously reported as causing MODY (red), present in large-scale sequencing datasets (green), dysfunctional protein in functional studies (purple), novel variants underlined.

3.3.8 Prevalence of HNF1A-MODY in Croatia

My study is the first to report subjects diagnosed with MODY in Croatia, where genetic testing for monogenic diabetes is not available. This means that individuals with MODY have been misdiagnosed as type 1 or type 2 diabetes and not being offered an optimal treatment.

I estimated the prevalence of HNF1A-MODY in Croatia based on sequencing of 477 individuals with C-peptide positive and β -cell antibody negative diabetes diagnosed

before 45 years of age. In the UK, the prevalence of HNF1A-MODY has been previously reported as 54 cases per 1 million of general population (Shields, Hicks et al. 2010).

Rare non-synonymous coding *HNF1A* variants were present in 13 individuals from the Croatian sample in my study and 9 (7 probands and 2 relatives) carried damaging/likely damaging *HNF1A* alleles. Therefore, 2.73% (95% CI 1.27 – 4.19%) of the Croatian study participants carried rare non-synonymous *HNF1A* alleles and 1.89% (95% CI 0.88 – 2.90%) had damaging/likely damaging *HNF1A* alleles (presumed HNF1A-MODY).

The Croatian Institute of Public Health records from 2015 stated that there were 260,092 subjects registered with diabetes in Croatia, 90% of whom had a diagnostic label of T2D (n=234,083), and 7.3% of whom had an onset of diabetes before the age of 45 years (n=17,088) (http://www.idb.hr/web_english/crodiab.htm; accessed October 12, 2016).

The most recent census from 2011 reported the population of Croatia as 4,284,889 (<https://www.dzs.hr/eng/censuses/census2011/censuslogo.htm>) which was estimated in 2015 at similar level.

Hence, this would suggest approximately 322 (95% CI 124 – 519) subjects with damaging/likely damaging *HNF1A* alleles (presumed HNF1A-MODY) in Croatia and a prevalence of 75 (95% CI 29 – 121) cases per million in the general population.

3.4 Discussion

In this chapter, I have shown that the systematic assessment of genetic variants is complex and requires multiple pieces of evidence for and against the disease causality of an allele

before the phenotypic effect can be assigned. Often an essential piece of evidence supporting the damaging effect of the allele is not available, such as co-segregation of the variant allele with diabetes in more than 2 family members. Another important feature supporting HNF1A-MODY is sensitivity to sulphonylureas which is often not recorded, especially with long duration of diabetes.

Sequencing *HNF1A* in a relatively unselected sample of individuals with diabetes found that 2.9% of them (30 probands of 1032 study subjects) with non-autoimmune diabetes diagnosed before 45 years of age carry rare non-synonymous *HNF1A* variants. In this study, 53% of rare non-synonymous *HNF1A* alleles were assigned damaging/likely damaging effect before functional data was included.

The proportion of study subjects harbouring damaging/likely damaging *HNF1A* variants of 1.94% (20 probands) is higher than the one reported by Flannick *et al.* (Flannick, Beer *et al.* 2013). That study aimed to characterise the spectrum of low-frequency variation (MAF<1%) in the main MODY genes in 4003 individuals from population based samples, which included subjects with (one third) and without diabetes (two thirds). The authors highlighted the limitations of previous databases of genetic variants reported to cause MODY (such as Human Gene Mutation Database, HGMD) by showing that 1.5-1.8% of the individuals from population-based samples carry these variants (including 12 variants in *HNF1A*). A smaller group of ~0.5% subjects in their study carried variants which were present in HGMD as causal for MODY but were also evolutionally conserved, predicted by bioinformatics as damaging and not present in 1000 Genomes project (including 5 *HNF1A*

variants). Those selection criteria would be fulfilled by damaging/likely damaging *HNF1A* alleles in my study and for that reason I compared proportion of these variants.

The large-scale sequencing dataset of GnomAD proved very useful for the variants c.-4A>G, p.E48K, p.T196A, p.G288W, p.G339S, p.H349Q, p.L389V and p.R583Q, all of which were relatively frequent for the rare alleles (MAF 0.06-0.007%) and also present in non-diabetic subjects from T2D-GENES project. Some of those variants are more frequent in selected ancestry populations and their allele frequency should always be reported for the appropriate ethnic group. For example, allele c.-4A>G is predominantly reported in South Asian subjects (189/200 alleles in GnomAD), variant p.L389V in Africans (158/169 alleles) and p.R583Q in European non-Finnish populations (108/125 alleles).

The spectrum of the functional prediction of *HNF1A* alleles found in this study and an overlap with *HNF1A* alleles associated with T2D is consistent with the existence of a phenotypic continuum of C-peptide positive β -cell antibody negative diabetes. This means that different *HNF1A* alleles, also influenced by other variants/genes, epigenetics or environmental factors, result in different allele penetrance and phenotypes ranging from MODY to typical T2D. Hence, finding an allele which can cause MODY should be interpreted in the context of clinical features to establish a diagnostic relevance.

Using the Sanger sequencing technique, visually reviewing all variants annotated and all sequencing with quality below 40 provides confidence for this study findings, especially since Sanger sequencing remains a gold standard technique still used in the clinical laboratories for confirmation of next-generation sequencing results. One limitation of my

study is not performing multiplex ligation-dependent probe amplification (MPLA) to look for whole exon/part of the gene deletions, with an exception of one patient with exon 1 deletion, who due to high clinical suspicion of MODY had MPLA done in the diagnostic laboratory. However, *HNF1A* exon deletions were reported in only 1.2% of the subjects with HNF1A-MODY (Colclough, Bellanne-Chantelot et al. 2013).

Based on sequencing *HNF1A* from 477 Croatian individuals with C-peptide positive and GAD antibody negative diabetes diagnosed before 45 years of age, I have estimated the prevalence of HNF1A-MODY in Croatia as 1.89% of diabetic subjects and 75 cases per million in the general population. The national register of all patients with diabetes provided accurate population data. This is similar to the prevalence of HNF1A-MODY reported by the Norwegian HUNT2 study of 1,972 diabetic patients which estimated a minimum prevalence of 0.4% of HNF1A-MODY, corresponding to 63 cases per million (Eide, Raeder et al. 2008). Similarly, a population study from Oxfordshire, UK reported 84 cases of HNF1A-MODY per million (Kropff, Selwood et al. 2011) and Shields *et al.* 108 cases of MODY (including 54 cases of HNF1A-MODY) per million in the general UK population (Shields, Hicks et al. 2010).

The advantage of my study compared to most studies reporting prevalence of MODY is the fact that all C-peptide positive and β -cell antibody negative participants diagnosed before 45 years of age had *HNF1A* sequencing, so fewer cases with *HNF1A* variants were missed by not narrowing the selection criteria for genetic testing. Limiting the age of onset of HNF1A-MODY to below 30 years would have missed 9 of 25 individuals (36%) with

damaging/likely damaging *HNF1A* alleles identified in this study. Similarly, using an Exeter MODY calculator as a screening tool would miss 52% of individuals with damaging/likely damaging *HNF1A* alleles in this study.

3.5 Conclusions

I have described 26 rare non-synonymous *HNF1A* variants found in 30 probands and 8 relatives. Around fifty percent of those alleles are likely to be causal for HNF1A-MODY based on a systematic assessment of all available evidence. The interpretation of a functional effect of the genetic variants is complex and must take into account all relevant aspects discussed in this chapter. Bearing in mind a functional spectrum of *HNF1A* alleles and multiple factors contributing the final phenotype, finding an allele which can cause MODY should be interpreted in the context of clinical features to establish a diagnostic relevance.

I have also estimated a prevalence of HNF1A-MODY in Croatia, which is similar to that previously reported in the UK, Norway and Germany.

Chapter 4

Functional studies of uncharacterised HNF1A protein variants

4.1 Introduction

In chapter 3, I discussed *HNF1A* variants identified during sequencing of *HNF1A* from 1032 individuals with young adult onset non-autoimmune diabetes from Croatia and the UK. There were 26 rare non-synonymous *HNF1A* variants identified, which were systematically assessed based on available evidence to assign a likely functional effect to each variant allele.

Guidelines for the interpretation of genetic variants recommend to validate the predicted effect of an allele experimentally using *in vitro* or *in vivo* models (MacArthur, Manolio et al. 2014, Kalia, Adelman et al. 2017). Functional characterisation is likely to be useful in contributing to the interpretation of an allele's phenotypic effect especially in the case of missense variants. However, caution is advised when extrapolating the results in cell lines *in vitro* to humans due to: the complexity of human biological pathways and interaction between genes/variants; differential expression of isoforms in different tissues; and developmental and environmental factors. Additionally, the function of a gene/protein variant assessed on the particular tissue-specific cell line may not be relevant to the particular disease. The experimental characterisation of *HNF1A* variants is discussed in more detail in section 4.4.

4.1.1 Experimental characterisation of rare non-synonymous *HNF1A* variants reported in literature

Seventeen of 26 rare non-synonymous *HNF1A* variants identified in this study had been experimentally characterised in the literature previously and are summarized in Table

4.1.1. Nine rare *HNF1A* variants (including three novel variants) lacked experimental data at the time of discovery.

Generally, functional characterisation of a transcription factor, such as HNF1A, includes: an assessment of transcriptional activity by a dual luciferase reporter system; protein expression of a transcription factor by western blotting; DNA binding by electrophoretic mobility shift assay (EMSA); and cytosol/nuclear localisation of the protein. Those methods have been commonly used in the functional assessment of HNF1A protein in the published literature. The first three methods were used in this study and were described in chapter 2, section 2.9.

The functional studies of *HNF1A* variant alleles reported to date had been performed using a variety of cell lines and promoters of genes for which HNF1A is a transcription factor (Table 4.1.1) (Vaxillaire et al. 1999, Galan, Garcia-Herrero et al. 2011, Najmi et al. 2016). For those studies, cell lysates for the assessment of HNF1A protein expression or DNA binding had been obtained from previously transfected and lysed cells (e.g. HeLa) (Hassanali, unpublished) or by TNT Coupled Reticulocyte Lysate System (Promega, Madison, USA) (Najmi, Aukrust et al. 2016). The latter method used rabbit reticulocytes and coupled transcription with translation in the tube to provide a higher yield of proteins. These differences in methodology make a direct comparison of the results difficult.

The function of the HNF1A protein variant had been shown to be reduced to <50% of the WT HNF1A protein in an assessment of transcriptional activity, protein expression and DNA binding in the case of 5 of 17 variants investigated (p.R229*, p.T260M, p.P291fs,

p.G292fs, p.P379fs; Table 4.1.1, Figure 4.1.1) (Yamagata, Oda et al. 1996, Vaxillaire, Abderrahmani et al. 1999, Yamada, Tomura et al. 1999, Bjorkhaug et al. 2003). Four of these are protein truncating variants (PTV) causing nonsense-mediated decay and haploinsufficiency (the exception is p.T260M) and all had been reported to co-segregate with young-onset diabetes in several families. Variant p.P379H had significantly reduced transcriptional activity (38-58% of the WT HNF1A) but normal HNF1A protein expression (Domínguez-López A 2005).

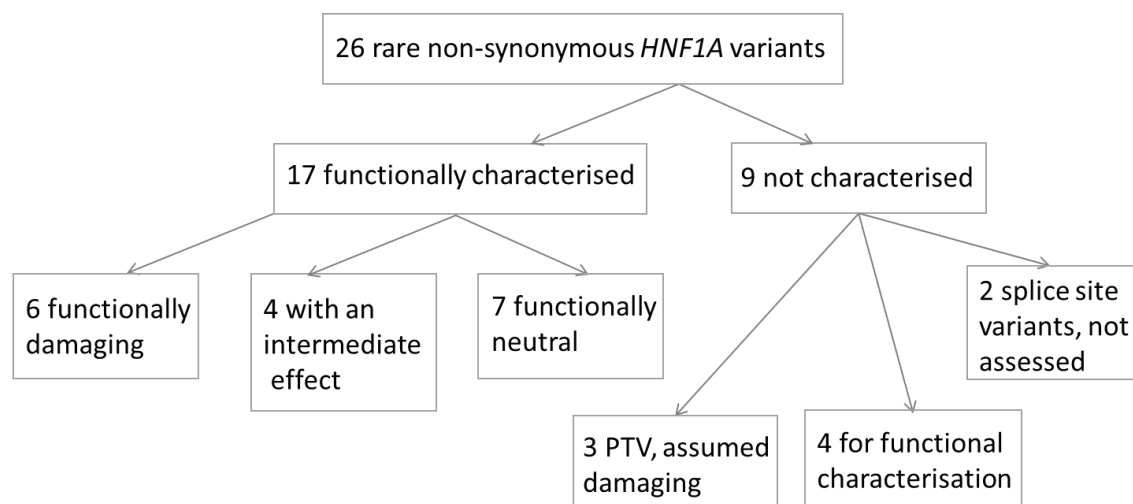


Figure 4.1.1 Rare non-synonymous *HNF1A* variants found in this study divided according to the functional characterisation results based on previous literature reports.

Seven of 17 *HNF1A* protein variants had previous experimental results comparable to the WT *HNF1A* protein (p.G47R, p.T196A, p.A251T, p.G339S, p.H349Q, p.P379R, p.G606S) (Galan, Garcia-Herrero et al. 2011, Balamurugan et al. 2016) (Hassanali, unpublished). Those *HNF1A* variants had undergone functional assessments of the transcriptional activity by the dual luciferase reporter system, protein expression by western blotting,

nuclear/cytosol localisation and/or DNA binding by EMSA, and all of those assays had been consistent with normal protein function (Table 4.1.1, Figure 4.1.1).

Variants p.E48K, p.L389V, p.T515M and p.R583Q had shown transcriptional activity of 60-80% of the WT HNF1A protein (Table 4.1.1), which suggests intermediate impact on the HNF1A protein function. The nuclear localisation of the above protein variants was similar to WT HNF1A protein (Najmi, Aukrust et al. 2016). DNA binding of these variants was not assessed, because the authors indicated that they were localised outside the DNA binding domain.

From the nine functionally uncharacterised rare *HNF1A* variants, three are PTVs (Exon 1 del, p.D135fs, p.L377fs) localized in exons 1-6 and likely to cause nonsense-mediated decay. These variant alleles had been reported to co-segregate with young adult-onset diabetes (Frayling et al. 2001, Ellard et al. 2007). I assumed these variants to be damaging to HNF1A protein function and they were not experimentally assessed in this study. A novel variant, p.S3C, was predicted to affect splicing so required an assessment of mRNA expression in fresh lymphocytes from the patient from Croatia, which was not possible to obtain. The 5'UTR variant, c.-4A>G, was relatively common (MAF 0.6% in South Asian

Coding DNA variant	Protein variant	Transcription activity [% of WT]	Protein expression [% of WT]	DNA binding [% of WT]	Nuclear localisation [% total]	Cell line and gene promoter used	Reference
c.-4A>G	n/a	n/a	n/a	n/a	n/a	n/a	n/a
c.1-326del	del exon 1	n/a	n/a	n/a	n/a	n/a	n/a
c.8C>G	S3C	n/a	n/a	n/a	n/a	n/a	n/a
c.139G>C	G47R	80-112%	110%	n/a	n/a	HeLa/INS1 cells with human alb/HNF4A promoter	Hassanali unpublished
c.142G>A	E48K	63%	n/a	n/a	78%	HeLa cells with rat alb promoter	Najmi 2016
c.404delA	D135fs	n/a	n/a	n/a	n/a	n/a	n/a
c.451G>A	G151S	n/a	n/a	n/a	n/a	n/a	n/a
c.586A>G	T196A	~100%	n/a	~100%	86%	MIN6/Cos7 cells with HNF4A/GLUT2 promoters, HeLa cells with rat alb promoter	Galan 2011, Najmi 2016
c.666G>T	K222N	n/a	n/a	n/a	n/a	n/a	n/a
c.685C>T	R229*	0-20%	n/a	11%	59%	MIN6 with GLUT2 promoter, Caco-2 with sucrose-isomaltase/DPP-4 promoter	Yamada 1999, Bjorkhaug 2003
c.751G>A	A251T	80-92%	98%	n/a	n/a	HeLa/INS1 cells with human alb/HNF4A promoter	Hassanali unpublished
c.779C>T	T260M	10-20%	10-20%	10-20%	n/a	HeLa cells with rat alb promoter	my own data
c.862G>T	G288W	n/a	n/a	n/a	n/a	n/a	n/a
c.871C>A	P291T	n/a	n/a	n/a	n/a	n/a	n/a
c.872delC	P291fs	0%	>100%	0%	80%	C33 cells with rat β -fibrinogen promoter	Vaxillaire 1999
c.872dupC	G292fs	0-10%	n/a [#]	n/a	n/a	HeLa with PKL promoter, C33 cells with rat β -fibrinogen promoter	Yamagata 1996
c.1015G>A	G339S	85-100%	~100%	n/a	~80%	HeLa with human Alb/CRP promoter	Hassanali unpublished
c.1047C>A	H349Q	160%	n/a	~100%	~80%	HeLa cells with rat alb promoter	Balamurugan 2016
c.1129delC	L377fs	n/a	n/a	n/a	n/a	n/a	n/a
c.1136_1137delCT	P379fs	6-62%	n/a	37%	<1%	HeLa with rat alb promoter, Caco-2 cells with sucrose-isomaltase promoter	Bjorkhaug 2003, Gu 2006
c.1136C>A	P379H	38-58%	~100%	n/a	n/a	HeLa and RINm5f cells with Ins promoter	Dominiguez-Lopez 2005
c.1136C>G	P379R	70-90%	n/a	~100%	n/a	MIN6/Cos7 cells with HNF4A/GLUT2 promoters	Galan 2011
c.1165T>G	L389V	70%	n/a	n/a	80%	HeLa cells with rat alb promoter	Najmi 2016
c.1544C>T	T515M	70-80%	95%	n/a	~80%	HeLa with human Alb/CRP promoter	Hassanali unpublished
c.1748G>A	R583Q	70%	n/a	n/a	80%	HeLa cells with rat alb promoter	Najmi 2016
c.1816G>A	G606S	105-125%	150-180%	n/a	60% (90% WT)	HeLa and INS1 cells with rat alb promoter	Bennett unpublished

Table 4.1.1 Functional characterisation of HNF1A protein variants based on previous literature reports. Variants uncharacterised experimentally before shaded in pale grey, results suggestive of significant reduction of protein variant function in bold; WT, wild-type; [#] mRNA expression 2% of the WT; alb, albumin.

ethnic group in GnomAD) and not included in functional studies as the systematic assessment, described in chapter 3, provided strong evidence that it is neutral to HNF1A protein function (Table 3.3.8).

Therefore, the four remaining uncharacterised *HNF1A* variants (p.G151S, p.K222N, p.G288W, p.P291T) and variant p.H349Q (also uncharacterised at the beginning of this study) were included in my functional characterisation experiments. Each of these variants is described in section 4.3.

4.1.2 Aims of the experiments

The objectives of the experiments described in this chapter were to:

1. Functionally characterise five *HNF1A* variants (p.G151S, p.K222N, p.G288W, p.P291T and H349Q), which had not been assessed at the time of starting the project;
2. Evaluate the contribution of functional data to understanding the clinical impact of selected *HNF1A* variants.

4.2 Methods

To functionally assess previously uncharacterised *HNF1A* variants, I used a pipeline for the functional assessment of HNF1A protein already established in Prof. Anna Gloyn's laboratory in OCDEM. This is summarised in Figure 4.2.1. I performed all functional work on the HeLa cell line (originally derived from human cervical cancer) (Scherer, Syverton et

al. 1953) as it does not express endogenous *HNF1A*. INS1 cells (originating from human insulinoma cells) express endogenous *HNF1A*, hence could also be used for the *HNF1A* variant assessment, having in mind that HNF1A-MODY in humans is caused by heterozygous inactivating *HNF1A* alleles.

Functional characterisation of several different HNF1A protein variants had been performed by my colleague, Dr Neelam Hassanali. Her work showed a similar performance in both cell lines used in inducing TA of human albumin by nine HNF1A protein variants tested (some also identified in my study: p.R203H, p.G47R, p.A251T and p.G339S) and two positive controls (Hassanali, unpublished). Therefore, I decided to use HeLa cells for all experiments. Maintenance of the HeLa cells was described in chapter 2, section 2.9.6.

I assessed the transcriptional activity using a dual luciferase reporter system, protein expression by western blotting and DNA binding by an electrophoretic mobility shift assay using HeLa cell lysates as described below. HNF1A is a transcription factor for more than 200 genes; therefore an assessment of transcriptional activity and a DNA binding of HNF1A protein were considered most important for the functional assessment of this protein. Cytosol/nuclear localization of the HNF1A protein was not assessed, based on the fact that all five tested variants had a transcriptional activity above 60% of the WT indicating that the amount of HNF1A protein was sufficient in the nucleus.

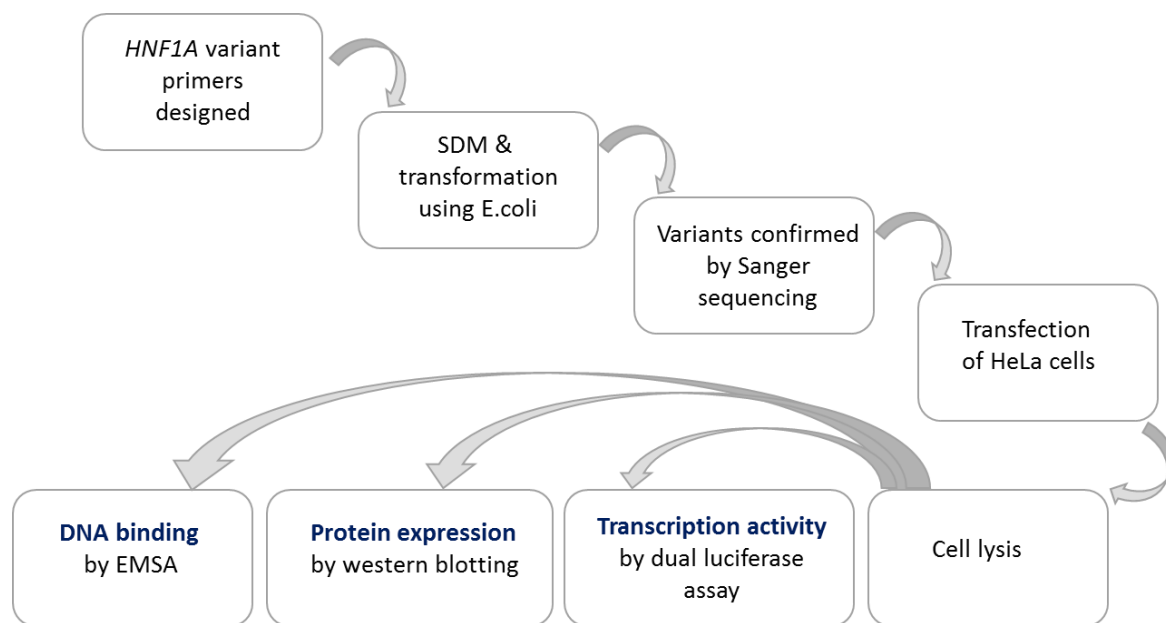


Figure 4.2.1 Work flow for the functional characterisation of the *HNF1A* variants; SDM, site-directed mutagenesis; EMSA, electrophoretic mobility shift assay.

4.2.1 Site-directed mutagenesis, transformation using *E.coli* and transfection of HeLa cells

cDNA of human *HNF1A* (NM_000545.5) was inserted into the pcDNA 3.1/His C plasmid including ampicillin resistance sequence (gift from KG Jebsen Center for Diabetes Research, University of Bergen) and used as a template in the mutagenesis of *HNF1A*. I designed forward and reverse primers for the five *HNF1A* variants investigated (p.G151S, p.K222N, p.G288W, p.P291T and H349Q, listed in Table 4.2.1) using the Quikchange Primer Design website (Agilent Technologies, Berkshire, UK). Site-directed mutagenesis (SDM) was performed to introduce specific variants into the cDNA of the *HNF1A* in the pcDNA3.1/His C plasmid as described in chapter 2, section 2.9.1.

DNA variant	Protein variant	Exon	WT sequence	Direction	Primers designed
c.451G>A	G151S	2	TGTCCCAACACCTCAACAAGGGCACTCCCATGAAGACGCAG	F	ccaacacccaacaagagcactcccatgaagac
				R	gtcttcattggagggtctctgttgagggtgttg
c.666G>T	K222N	3	CAGGCCTATGAGAGGCAGAAGAACCCTAGCAAGGAGGAGCG	F	tccaggcctatgagaggcagaataaacctagcaag
				R	cttgctagggtattctgcctctcataggcctgga
c.862G>T	G288W	4	TGGCCATGGACACGTACAGCGGGCCCCCCCAGGGCCAGGC	F	gacacgtacagctggccccccag
				R	ctggggggggccagctgtacgtgtc
c.871C>A	P291T	4	ACACGTACAGCGGGCCCCCCCAGGGCCAGGCCGGGACCT	F	cggggccccccacaggggccagg
				R	cctggccctgtggggggcccg
c.1047C>A	H349Q	5	ACAGTGTCTACACCCCTCCAACAAGTGTCCCCACGGGCCT	F	tctacaccctccaacaagtgtccccac
				R	gtgggggacactgttgagggtgtaga

Table 4.2.1 Sequence of the *HNF1A* variants for the functional characterisation and the designed specific primers for site-directed mutagenesis; F, forward; R, reverse. Wild-type nucleotide marked in grey, nucleotide substitution in primers marked in red.

The XL1-blue strain of *E.coli* was transformed with mutagenized plasmids and grown to produce a sufficient amount of mutagenized plasmid cDNA (mini-prep). An aliquot of the *E.coli* cell suspension was mixed with 7% DMSO to generate a DMSO stock and frozen at -80°C (all described in chapter 2, section 2.9.2). Subsequently, cDNA of the mutagenized plasmids was extracted according to the description in section 2.9.3 and confirmed by sequencing using T7 forward primer located upstream to *HNF1A* and designed *HNF1A* cDNA forward primers F2, F3, F4, F6 (Eurofins Genomics, Wolverhampton, UK) to confirm the presence of the desired nucleotide change and ensure that no additional nucleotide substitutions in the remaining part of the *HNF1A* had been introduced. The sequences of the primers were listed in Table 2.9.1.

When the successful mutagenesis of each *HNF1A* variant was confirmed, a scraping from the DMSO stock of *E.coli* suspension was used to generate a larger stock of the mutagenized plasmids which were stored at -20°C as midi-prep and used for all experiments.

HeLa cells were transiently transfected with plasmids containing mutagenized cDNA using Lipofectamine 2000 (Invitrogen, Thermo Fisher Scientific, Waltham, MA, USA) as described in chapter 2, section 2.9.8.

To produce cell lysates for the later assessment of transcriptional activity, HeLa cells were transfected with plasmid mixes, which included plasmid containing cDNA of the *HNF1A* (WT or mutagenized), plasmid containing cDNA of rat albumin promoter and firefly luciferase (pGL3-RA-FL reporter), plasmid containing cDNA of Renilla luciferase (pRL-SV40) as well as empty vector. The amount of each pcDNA is given in chapter 2, Table 2.9.2.

4.2.2 Transcriptional activity using dual luciferase reporter system

I assessed the transcriptional activity (TA) of the variants of HNF1A protein using a dual luciferase reporter system with rat albumin promoter as a reporter construct. The albumin promoter was used because it has the strongest consensus binding site for HNF1A protein. Firefly luciferase reporter (cloned into pGL3 vector) was utilised as a measure of transcriptional activity of the wild-type or mutagenized HNF1A protein. *Renilla* luciferase reporter was used as an internal control (inserted into pRL-SV40). The measurement of luminescence of both luciferases after the injection of their substrates (LARII and “Stop & Glow”, Promega, Southampton, UK) was performed on a Multimode Plate Reader, EnSpire (Perkin Elmer, Waltham, USA) as explained in Chapter 2, section 2.9.9. Relative luciferase units for Firefly luciferase were normalised against *Renilla* luciferase units and presented as means ($\pm$ SEM, standard error of means) of three

technical replicates per HNF1A protein variant. Subsequently, means ($\pm$ SEM) from three biological replicates were calculated and presented as mean percentage of the WT HNF1A protein activity.

4.2.2.1 Comparison of transcriptional activity using different midi-preparations of the HNF1A variant

First, I set up an experiment to compare different preparations (midi-preps) of the same *HNF1A* variant pcDNA when both preparations originated from the same DMSO stock but were made in different months. The older midi-preps have been frozen and thawed more often, which could affect the quality of the variant stored samples. I have also compared variant preparations made by two different researchers (myself and Dr Amanda Bennett). For that experiment I chose one variant used as a MODY positive control (p.T260M) and one variant known to intermediately reduce the transcriptional activity of HNF1A (p.E508K). I also checked the performance of both variants when the pcDNA midi-preparation was diluted by 10-fold as small volumes are likely to be less accurate.

Using different midi-preps of the HNF1A protein variant p.T260M resulted in reproducible transcriptional activity with standard deviation (SD) of 1.4%. The difference in transactivation of p.E508K was bigger when two different preparations [preparation done by myself (A) or Dr Amanda Bennett (B)] were used (SD 22.8%, Figure 4.2.2). Nonetheless, the transcriptional activity of variant p.E508K was reproducible when the same preparation of this variant was used again in subsequent experiments.

Diluting both variant pcDNAs before transfection gave similar TA results when compared to undiluted preparations (difference in the SD ranged 1.2-15%, Figure 4.2.2).

Based on the above, I used the same midi-preparation of each variant for all experiments related to the particular variant (TA, WB and EMSA). All tested variant preparations used for transfection were 10-fold diluted.

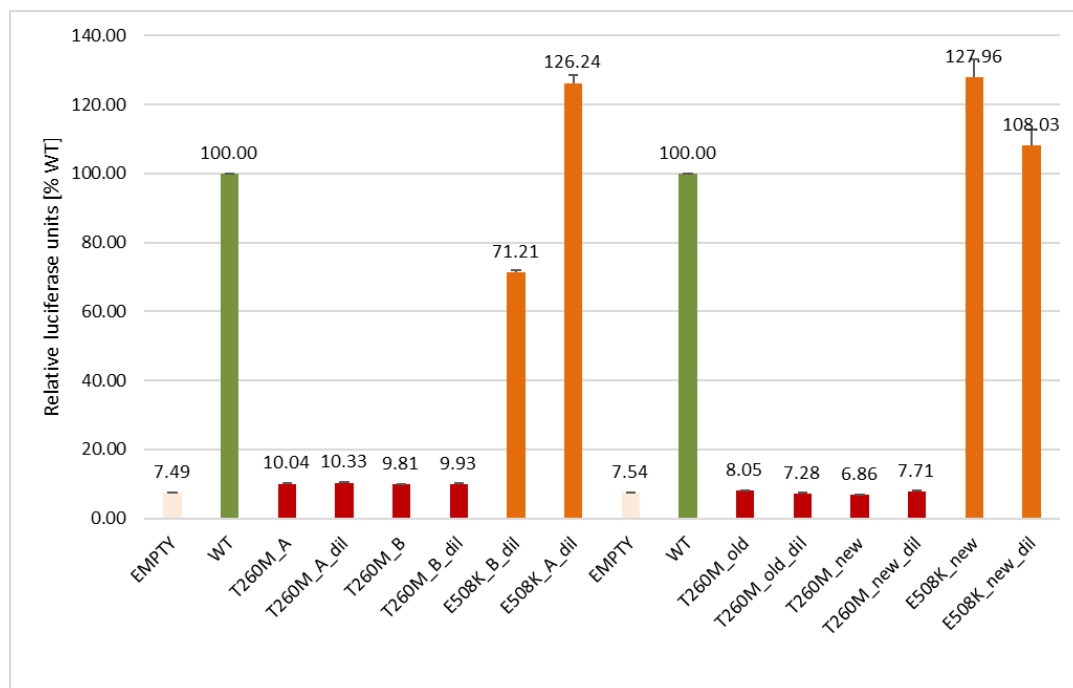


Figure 4.2.2 Transcription activity of different pcDNA midi-preparations of variant p.T260M and p.E508K extracted by different researchers (A or B) and using an old or new variant preparation, as well as diluted (dil) vs undiluted pcDNA preparations. The bars represent relative luciferase units presented as a mean percentage ($\pm$ SEM) of the WT HNF1A from 3 technical replicates.

4.2.3 HNF1A protein expression by chemiluminescent western blotting

Twenty four hours post-transfection, HeLa cells were lysed for western blotting and DNA binding assessment as described in chapter 2, section 2.9.10. The total protein was

quantified with Bradford assay (Bio-Rad Laboratories, Watford, UK) using VersaMax microplate reader and Softmax Pro v5.2 (Molecular Devices, Sunnyvale, CA, USA) as described in section 2.9.11.

I assessed protein expression by chemiluminescent western blot (WB) utilising anti-HNF1A antibodies and β -tubulin as an internal control (anti-HNF1A sheep polyclonal antibody and anti- β -tubulin rabbit antibody, Santa Cruz Biotechnology, USA). Total protein (10 μ g/variant) was mixed with Laemmli buffer and DTT and denatured at 80°C, then loaded onto a 4-20% Criterion TGX Stain-free Precast gel (1mm, 26 wells, Bio-Rad, Hercules, USA). The Protein All Blue standard (Bio-Rad, Munich, Germany) was applied as a reference protein ladder. The gel was activated using the Chemi-Doc MP Imaging System (Bio-Rad, Watford, UK) and proteins transferred to a PVDF Midi membrane using Trans-Blot Turbo (Bio-Rad, Watford, UK). Consequently, the membrane was cut to separate HNF1A and β -tubulin bands based on their molecular weight and blocked first with primary then secondary antibodies, exposed to Clarity Western ECL (Bio-Rad, Watford, UK) and imaged with Chemi-Doc MP Imaging System. All steps of Western blot are explained in detail in chapter 2, section 2.9.12.

The results from WB were presented as mean ($\pm$ SEM, standard error of means) percentages of the WT HNF1A protein of four biological replicates per HNF1A protein variant.

4.2.4 DNA binding using Electrophoretic mobility shift assay

I used the Electrophoretic Mobility Shift Assay (EMSA) to assess binding of the WT or variant HNF1A protein to a Cy5-labelled double-stranded oligonucleotide of the rat albumin promoter representing a part of the HNF1A recognition sequence. The total protein (10 µg) of the WT or variant HNF1A was mixed with the binding mix (Odyssey Infrared EMSA kit, LiCor, Lincoln, USA) and the Cy5-labelled oligonucleotide probe, incubated and loaded onto the 5% Criterion Precast gel (Bio-Rad, Hercules, USA) as defined in section 2.9.13. Next, the post-electrophoresis gel was visualised on the ChemiDoc MP Imaging System with near-infrared filter. Adding anti-HNF1A antibodies (anti-HNF1A sheep polyclonal antibody, Santa Cruz Biotechnology, USA) to a 2nd sample of WT HNF1A protein demonstrated a supershift effect and confirmed the localisation of the HNF1A band on the gel. EMSA is described in detail in chapter 2, section 2.9.13.

The results from EMSA were presented as mean ($\pm$ SEM, standard error of means) percentages of the WT HNF1A protein activity of four biological replicates per HNF1A protein variant. Cell lysates contained the same amount of total protein including HNF1A for each variant. Therefore, to correct for the HNF1A protein amount in each variant preparation, the mean percentage of the WT of each variant band from EMSA was divided by the mean percentage of the WT from western blot for the same four biological replicates.

4.2.5 Statistical analysis

Results from each step of the functional studies are presented as mean ($\pm$ SEM, standard error of means) percentage of the WT HNF1A protein and compared using SPSS v23 and one-way analysis of variance (ANOVA). A pairwise comparison of mean percentage of the WT HNF1A protein for the investigated variant was used, and the p value was adjusted for multiple comparisons [the Least Significant Difference (LSD) test]. I chose to use one-way ANOVA for the comparison of HNF1A protein variants function to capture differences in the variance of all HNF1A protein variants as they all went through the same experimental steps.

4.3 Characteristics of HNF1A protein variants undergoing functional assessment

Ten HNF1A protein variants (5 tested and 5 control variants) underwent functional assessment: two novel variants (p.G151S, p.K222N), p.G288W, p.P291T and p.H349Q, which had not been functionally characterised at the time of starting the project, and five control variants: p.P112L, p.T260M, p.R203H, p.H179H and p.E508K.

The positions of tested and control variants in HNF1A protein are shown in Figure 4.3.1.

Each experiment included empty plasmid, wild-type (WT) *HNF1A*, two or three positive controls of the *HNF1A* variants from DNA binding domain (p.P112L, p.T260M, p.R203H) which had been previously functionally characterised, one synonymous variant used as a negative control (p.H179H) and one variant associated with an increased risk of T2D (p.E508K). The selection of three positive control *HNF1A* variants (p.P112L, p.T260M,

p.R203H) was based on established evidence for causing MODY such as co-segregation of the variant allele with the young adult-onset diabetes in multiple families and supporting functional data in the literature or in the work from our group. The synonymous variant p.H179H was used as a negative control and expected to behave as WT HNF1A protein. Variant p.E508K had previous functional data available with intermediate results.

Each experiment was repeated three times on further passages of HeLa cells to obtain four biological replicates for WB and EMSA, while transcription activity assessment had three technical replicates in each of three biological replicate.

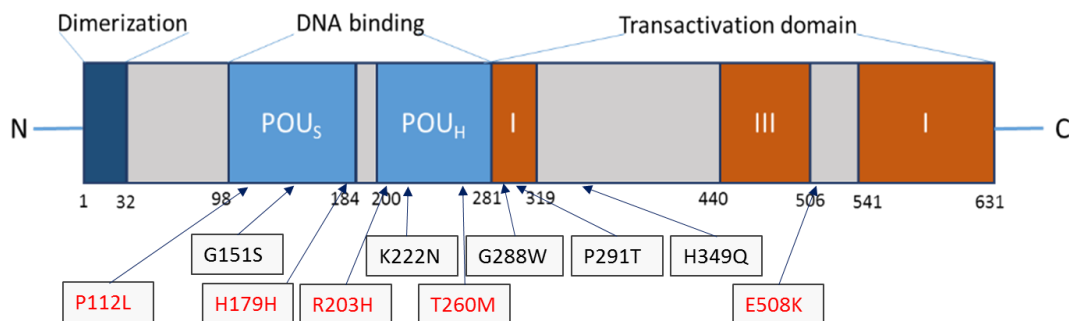


Figure 4.3.1 *HNF1A* variants investigated in this study with location of the codon they affect in the *HNF1A* protein; variants tested shown in the black font, control variants in the red font.

4.3.1 Features of *HNF1A* variants undergoing functional characterisation

Variant c.451G>A, p.G151S (Figure 4.3.2) is a novel *HNF1A* variant and thus had not been reported in the literature previously. It is localised in exon 2, within the DNA binding domain, and therefore affecting all three isoforms of the *HNF1A* protein. The Croatian individual carrying the p.G151S variant, discovered in this study, was diagnosed with diabetes at 35 years of age and was initially only treated with diet. Ten years after

diagnosis of diabetes, she had a BMI of 26.3kg/m², was treated with metformin, glimepiride (SU) and vildagliptin with a HbA1c of 6.3% and her hsCRP was 0.8mg/L. The effect of adding sulphonylurea is unknown in this case so cannot support disease causality of this allele. Her mother and maternal grandmother have diabetes but were not available for sequencing. All three bioinformatic tools used (SIFT, Polyphen 2 and Provean), predicted allele c.451G>A, p.G151S to be functionally damaging and this codon is conserved throughout evolution (the same in 7/7 species).

Variant c.666G>T, p.K222N (Figure 4.3.3) is also novel and localised in exon 3, within the DNA binding domain. The British individual carrying this variant was diagnosed with diabetes at the age of 18 and her diabetes was controlled with diet for 5 years, at which point a low dose of sulphonylurea (20mg daily) was added. This resulted in improved glycaemia with no episodes of hypoglycaemia. Her BMI five years from diagnosis of diabetes was 32kg/m² and her CRP was 0.2mg/L. She has a strong family history of diabetes with both parents, both grandmothers and maternal grandfather having diabetes (Figure 4.3.3). Her mother is also obese but does not have the same *HNF1A* allele, while her father and his mother were not available for genetic testing. Allele c.666G>T, p.K222N was predicted by all *in silico* tools to be damaging to HNF1A protein function and this codon is conserved throughout evolution (the same in 7/7 species).

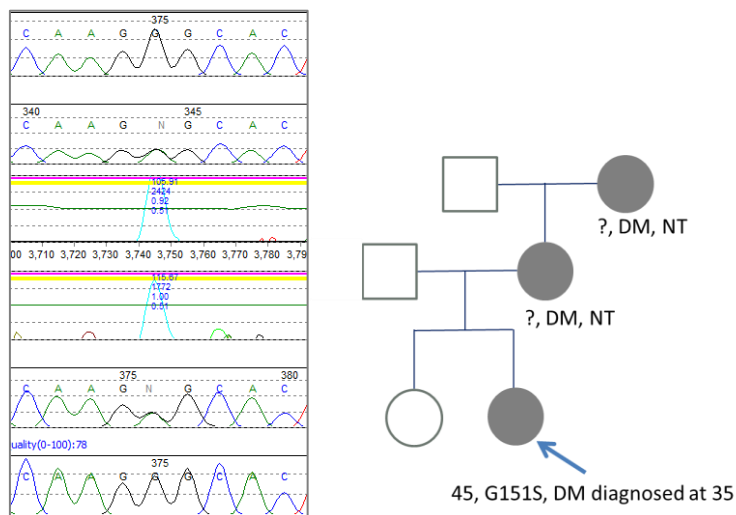


Figure 4.3.2 Electropherogram from capillary sequencing showing HNF1A allele c.451G>A, p.G151S (on the left) and family tree of the proband (on the right); individuals with diabetes shaded in grey; square represents man and circle woman; arrow marks the proband; DM, diabetes mellitus; NT, not tested.

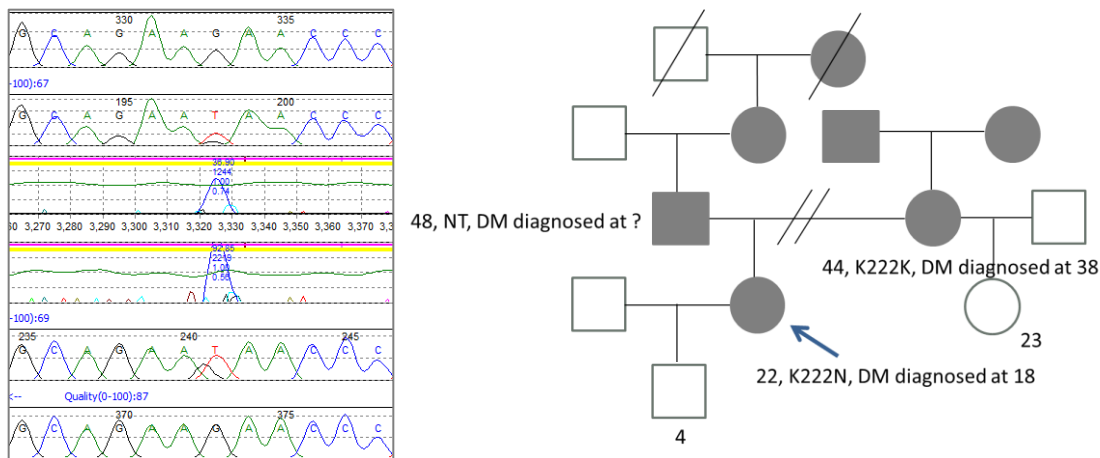


Figure 4.3.3 Electropherogram from capillary sequencing identifying HNF1A allele c.666G>T, p.K222N (on the left) and family tree of the proband (on the right); individuals with diabetes shaded in grey; square represents man and circle woman; line across square or circle indicates deceased person; //, divorced; arrow marks the proband; DM, diabetes mellitus; NT, not tested.

Variant c.862G>T, p.G288W (Figure 4.3.4) was reported in GnomAD in 18/264,764 alleles (MAF 0.007%, in 15 non-Finish European and 3 Ashkenasi Jewish individuals). A

synonymous variant in this codon (c.864G>C, p.G288G) has a reported MAF of 27% in GnomAD. The Croatian subject with variant p.G288W was diagnosed with diabetes at the age of 39, his BMI was 38.9kg/m² two years from diagnosis and he was diet controlled with HbA1c of 5.5%. His CRP was 15.7mg/L. His mother carries the same variant but at age 66 years she does not have diabetes despite being obese (BMI 30.9kg/m²). This allele is predicted by SIFT and Provean to be neutral to the protein function, while Polyphen 2 predicts it as likely damaging. The wild-type amino-acid in this position is the same in 5/8 species tested.

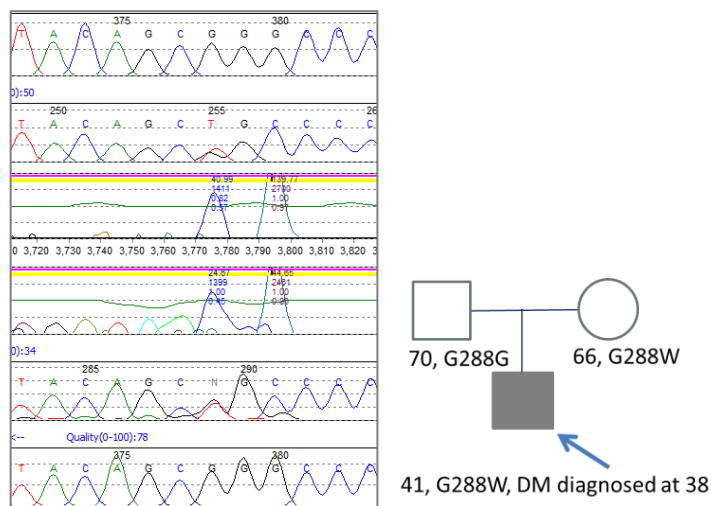


Figure 4.3.4 Electropherogram from capillary sequencing identifying 2 *HNF1A* alleles: rare c.862G>T, p.G288W and common c.864G>C, p.G288G (on the left) and the family tree of the proband (on the right); individuals with diabetes shaded in grey; square represents man and circle woman; arrow marks the proband; DM, diabetes mellitus.

Variant c.871C>A, p.P291T is located in exon 4 at the beginning of the transactivation domain (Figure 4.3.1) and lies in the poly-C region commonly affected by substitutions.

Two variants were captured in this region in the proband: common synonymous variant p.G288G and p.P291T (Figure 4.3.5). Allele c.871C>A, p.P291T had been previously reported as MODY-causing (Bellanne-Chantelot, Carette et al. 2008, Thanabalasingham, Shah et al. 2011) including co-segregation of the allele with diabetes in the family (Thanabalasingham, Shah et al. 2011). It was also reported in 2/234,764 alleles in GnomAD with MAF of 0.0008% and none of non-diabetic subjects in T2D-GENES project carried this allele. However, p.P291T is predicted by all three *in silico* tools to be neutral. In the present study, all three British family members carrying this *HNF1A* allele were obese (BMI 31-38.8kg/m²) with diabetes diagnosed at 27, 33 and 63 years of age. The mother of the proband was obese at diagnosis, subsequently lost 6kg in weight and had remained diet-controlled for 11 years with current HbA1c 5.8%, despite a BMI of 31kg/m². The proband and his cousin were not sensitive to sulphonylureas and quickly progressed to other treatment. The CRP in carriers of this allele was 0.7, 6.8 and 14.7mg/L.

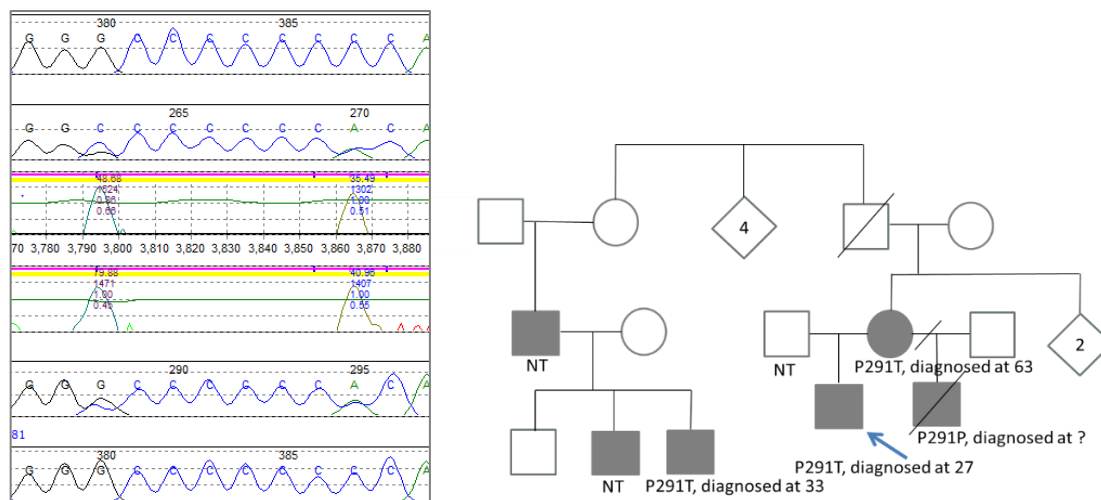


Figure 4.3.5 Electropherogram from capillary sequencing identifying 2 *HNF1A* alleles: common c.864G>C, p.G288G and rare c.871C>A, p.P291T (on the left) and the family tree of the proband

(on the right); individuals with diabetes shaded in grey; square represents man and circle woman; the number in the rhomboid shape gives a total number of children with gender not specified; arrow marks the proband; NT, not tested.

Variant c.1047C>A, p.H349Q (Figure 4.3.6) is in exon 5, within the transactivation domain and this allele had been reported in 16 heterozygous individuals in GnomAD, all of South Asian ethnicity (16/246,212; MAF 0.006%). This variant allele was also present in three non-diabetic Asian individuals in the T2D-GENES project and was predicted by all three bioinformatics tools as neutral to HNF1A protein function. The individual carrying this variant in the current study was also South Asian, was diagnosed with diabetes at age 33 years and is treated with insulin 13 years from diagnosis, as oral anti-glycaemic agents were not sufficient to control his hyperglycaemia. Both his parents have diabetes, as does his sibling and child but none of them was available for genetic testing (Figure 4.3.6). His CRP was 1.01mg/L. The functional characterisation of variant c.1047C>A, p.H349Q had been published at the end of my project and nuclear localisation and DNA binding of this HNF1A protein variant were both similar to the WT HNF1A protein and transcriptional activity was 160% of the WT HNF1A (Balamurugan, Bjorkhaug et al. 2016).

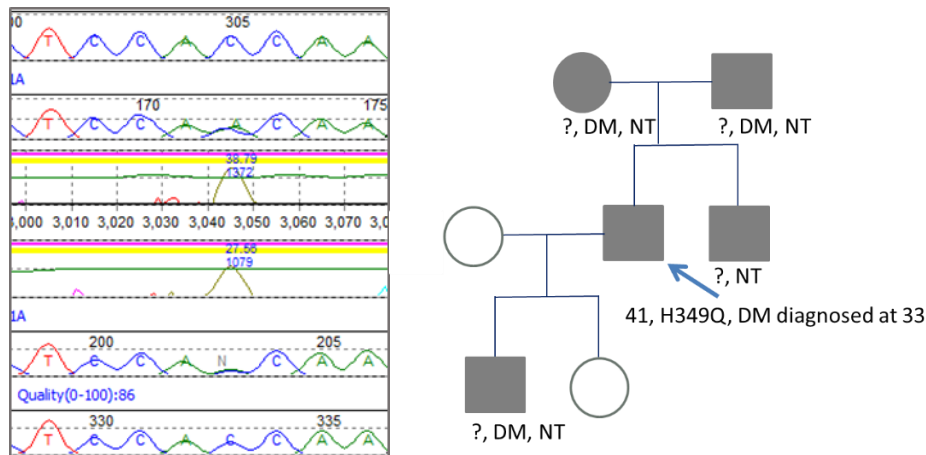


Figure 4.3.6 Electropherogram from capillary sequencing showing *HNF1A* allele *c.1047C>A*, *p.H349Q* (on the left) and family tree of the proband (on the right); individuals with diabetes shaded in grey; square represents man and circle woman; arrow marks the proband; DM, diabetes mellitus; NT, not tested.

4.3.2 *HNF1A* variants used as controls

HNF1A variant alleles (*c.335C>T*, *p.P112L*; *c.608G>A*, *p.R203H*; *c.779C>T*, *p.T260M*) are known to cause *HNF1A*-MODY and had been reported by 2013 in 13-19 MODY families each (Colclough, Bellanne-Chantelot et al. 2013). Co-segregation of the *HNF1A* allele with diabetes had been demonstrated for *p.P112L* and *p.T260M* (Glucksmann et al. 1997, Bjorkhaug et al. 2000), while sequencing of family members of the probands with *p.R203H* had been incomplete but inheritance of diabetes was consistent with an autosomal dominant pattern (Awata et al. 1998, Ng et al. 1999). All three variants used as positive controls are in the DNA binding domain and were predicted by all three bioinformatic tools as damaging to protein function.

Variant *c.335C>T*, *p.P112L* had been reported in 14 MODY families by 2013 (Colclough, Bellanne-Chantelot et al. 2013) and functionally characterised previously using HeLa cells

and rat albumin promoter (Bjorkhaug, Ye et al. 2000). In this study, variant p.P112L had a transcriptional activity (TA) of 18% of the WT HNF1A, significantly reduced DNA binding performed by EMSA (30% of the WT) and nuclear localisation similar to the WT HNF1A protein.

Variant c.608G>A, p.R203H had been reported in 19 MODY families by 2013 (Colclough, Bellanne-Chantelot et al. 2013). This variant had been analysed by my colleague from Prof. Anna Gloyn's group, Dr Neelam Hassanali, who showed that transcriptional activity of this variant was reduced to 60% of that of the WT HNF1A protein. That work had been performed using HeLa cells with the human albumin promoter and with INS1 cells using the HNF4A P2 promoter. On the other hand, p.R203H had significantly reduced transcriptional activity (20% of the WT) when HeLa cells with the HNF4A P2 promoter were used (Hassanali, unpublished). Protein expression of p.R203H using western blot was reduced to 70% of the WT, but this was not statistically significant.

Variant c.779C>T, p.T260M had been reported in 13 MODY families by 2013 and co-segregation of this allele with young onset-diabetes in a large family had been published (Glucksmann, Lehto et al. 1997). In the present study one proband, together with his sister and father, carried p.T260M. All three family members had young adult-onset diabetes, were slim and sensitive to a low dose of sulphonylurea. Although a search of the literature did not identify any functional data on p.T260M, unpublished results from Dr Amanda Bennett from Prof. Anna Gloyn's group were consistent with significantly impaired HNF1A protein function of this variant.

Variant c.1522G>A, p.E508K was used as an example of a non-synonymous *HNF1A* allele associated with an increased risk of type 2 diabetes in Latino populations [OR 5.48 (95% CI 2.83-10.61)] and its functional results had been reported previously (Estrada, Aukrust et al. 2014). *In vitro* characterisation of this variant using HeLa cells and rat albumin promoter had reported transcriptional activity reduced to 39% of the WT HNF1A and normal DNA binding of the protein (Estrada, Aukrust et al. 2014).

Finally, I used a synonymous variant c.537T>C, p.H179H as a negative control expecting that it will generate functional results similar to the WT HNF1A protein. Thus, it was used as a technical control for the WT HNF1A to assure that similar general experimental conditions have been achieved in the case of all variants tested.

4.4 Results

4.4.2 Transcriptional activity (TA) of the HNF1A protein variants using the dual luciferase reporter system

The mean normalized transcriptional activity (TA) of variant p.G151S was 82.8 (SEM $\pm$ 8.9) %, p.K222N 76.1 ($\pm$ 7.0) %, p.G288W 73.3 ($\pm$ 8.0) % and p.P291T 59.2 ($\pm$ 7.0) % of the WT HNF1A (Table 4.4.1, Figure 4.4.6). The reduction was statistically significant for variant p.G288W (p=0.04, one-way ANOVA) and p.P291T (p=0.001). The transactivation of p.H349Q was similar to that of the WT HNF1A (112.5 ($\pm$ 6.1) % WT) and consistent with recently published results (Balamurugan, Bjorkhaug et al. 2016).

The results from three biological replicates (each including three technical replicates, hence $n=9$ for each variant) were reproducible for each variant with a reasonably narrow standard error of means (0.3-14.5%) and standard deviation of 0.4-20.5 % WT.

The synonymous variant p.H179H had a transcriptional activity of 120.5 (± 20.1) % of the WT but the difference in normalized transcriptional activity of this variant was not significant ($p=0.09$). This difference in SEM of $\leq 20\%$ of the WT as well as SD of $\leq 20\%$ of the WT in the synonymous variant suggests that transcriptional activity of any variant in the range of 80-120% of the WT is likely to be non-significant.

The control variants p.T260M and p.R203H performed similarly to the unpublished results from our group (Bennett, Hassanali, unpublished) (Table 4.4.1). The mean normalized TA of p.T260M was 7.8 (± 1.1) % of the WT HNF1A in my experiments and 7.0 (± 1.0) % in Dr Bennett's experiments (Bennett, unpublished). Mean TA of variant p.R203H was 77.2 (± 9.5) % of the WT in my study, consistent with previous results from our group (60.0 (± 5.0) % of the WT) using similar experimental conditions (Hassanali, unpublished), with the exception of human rather than rat albumin promoter. An assessment of TA did not include variant p.P112L because it was added later as a third positive control for the assessment of protein expression and DNA binding.

Variant	n1	n2	n3	mean (n1-3) [%WT]	SD	SEM	P value*
Empty	4.4	7.3	9.6	7.1	2.6	1.8	1.5×10^{-7}
WT	99.3	100.0	100.0	99.8	0.4	0.3	n/a
T260M	6.5	7.6	9.5	7.8	1.5	1.1	1.6×10^{-7}
R203H	66.8	79.4	85.4	77.2	9.5	6.7	0.073
E508K	53.5	70.8	58.4	60.9	8.9	6.3	0.035
G151S	71.0	81.3	96.1	82.8	12.6	8.9	0.172
K222N	67.7	73.7	87.0	76.1	9.9	7.0	0.066
G288W	60.2	80.0	79.6	73.3	11.3	8.0	0.040
P291T	62.3	48.2	67.2	59.2	9.9	7.0	0.001
H349Q	102.5	117.8	117.1	112.5	8.6	6.1	0.329
H179H	143.0	104.7	113.7	120.5	20.1	14.2	0.095

Table 4.4.1 Mean transcriptional activity (TA), SD (standard deviation) and SEM (standard error of means) of WT and variant HNF1A using a dual luciferase reporter system in HeLa cells and rat albumin promoter (n=3 biological replicates). Results are expressed as a percentage of WT HNF1A; *p value given for the pairwise comparison by one-way ANOVA, significant p<0.05 in bold.

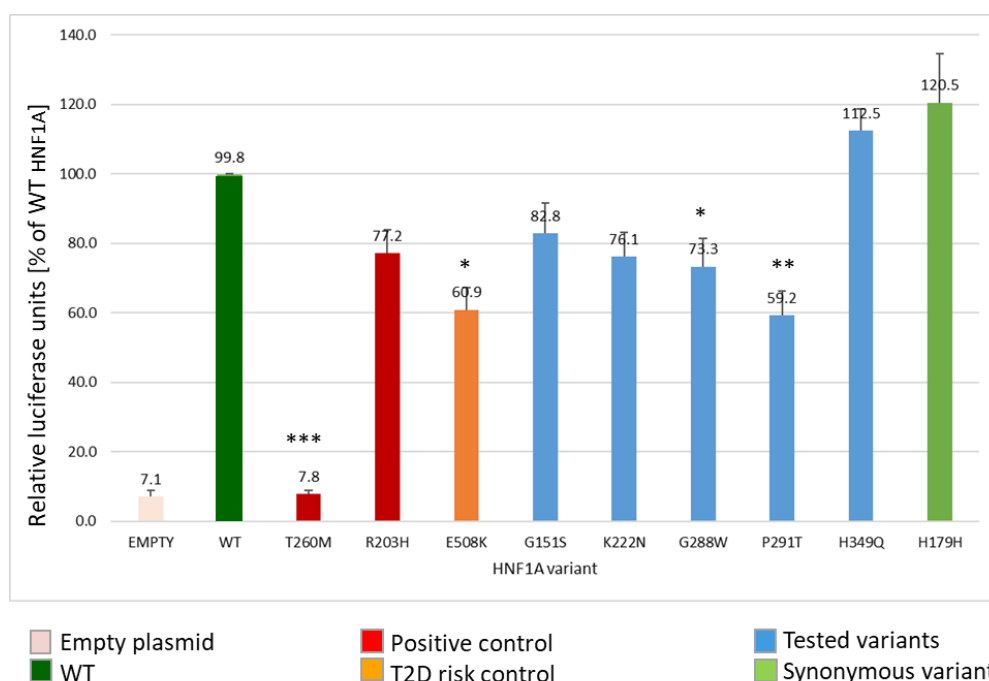


Figure 4.4.1 Transcriptional activity (TA) of WT and variant HNF1A using a dual luciferase reporter system in HeLa cells and rat albumin promoter (n=3 biological replicates). Variants p.T260M and p.R203H were used as positive controls, p.E508K as variant associated with T2D, p.H179H as negative control. Bars represent mean relative luciferase units expressed as % of the WT with

standard error of means; Comparison between variants and WT HNF1A was made using one-way ANOVA; *** $p < 0.0001$, ** $p < 0.01$, * $p < 0.05$.

The mean transactivation induced by p.E508K in my experiments was 60.9 (± 6.3) % of the WT HNF1A versus published previously 39.0% of the WT (Estrada, Aukrust et al. 2014). Although this difference could be within the limits of experimental error ($\pm 20\%$ in my experiments), the result of Estrada *et al.* showed a more significant reduction in TA, while my result represent a mild reduction of TA when compared to the WT HNF1A ($p < 0.0001$ vs $p < 0.05$).

As the variant p.E508K is associated with T2D rather than causing the MODY phenotype, results from my experiments better correlate with the phenotype of the individuals carrying that *HNF1A* allele. Despite the same plasmid constructs and HeLa cells with rat albumin promoter being used in both laboratories, differences in both protocols exist and may account for the final difference in the results. One of them is the method for obtaining cell lysates for the assessment of HNF1A protein. I obtained my protein lysates from previously transfected HeLa cells while Estrada *et al.* obtained cell lysates by TNT Coupled Reticulocyte Lysate System (Promega, Madison, USA).

4.4.3 HNF1A protein expression

The HNF1A protein expression of variants p.G151S, p.G288W, p.P291T and p.H349Q examined by Western blotting was similar to the WT HNF1A protein [102.8-136.3 (± 3.3 -18.5) % of WT, all $p > 0.05$, Table 4.4.2]. Variant p.K222N had significantly increased transcription activity [145.5 (± 14.7) % of the WT, $p = 0.023$, Table 4.4.2, Figure 4.4.7].

A mean normalized HNF1A protein expression was only significantly reduced in two positive controls, p.P112L [40.5 ($\pm$ 5.8) % of the WT, $p=0.004$] and p.T260M [40.3 ($\pm$ 8.2) % of the WT, $p=0.004$].

Variant	n1	n2	n3	n4	mean	SD	SEM	p value
Empty	2.0	3.0	3.0	8.0	4.0	2.9	2.0	1.6×10^{-5}
WT	100.0	100.0	100.0	100.0	100.0	0.0	0.0	n/a
P112L	45.0	37.0	32.0	48.0	40.5	8.2	5.8	0.004
R203H	132.0	61.0	109.0	95.0	99.3	24.7	17.5	0.969
T260M	43.0	50.0	41.0	27.0	40.3	11.6	8.2	0.004
G151S	175.0	110.0	117.0	108.0	127.5	4.7	3.3	0.158
K222N	203.0	103.0	143.0	133.0	145.5	20.8	14.7	0.023
G288W	158.0	122.0	107.0	158.0	136.3	26.2	18.5	0.066
P291T	124.0	89.0	103.0	137.0	113.3	24.7	17.5	0.492
H349Q	161.0	64.0	92.0	94.0	102.8	16.8	11.9	0.886
H179H	152.0	74.0	104.0	n/a	123.3	36.3	25.7	0.631
E508K	n/a	110.0	97.0	67.0	91.3	22.1	15.6	0.679

Table 4.4.2 Protein expression of WT and variants HNF1A using densitometry from western blotting presented as a mean ($\pm$ SEM) percentage of the wild-type (WT) from four biological replicates. All results normalized for B-tubulin to assure equal amount of protein lysate has been loaded. One-way ANOVA was used to compare means; significant $p<0.05$; n/a, not available.

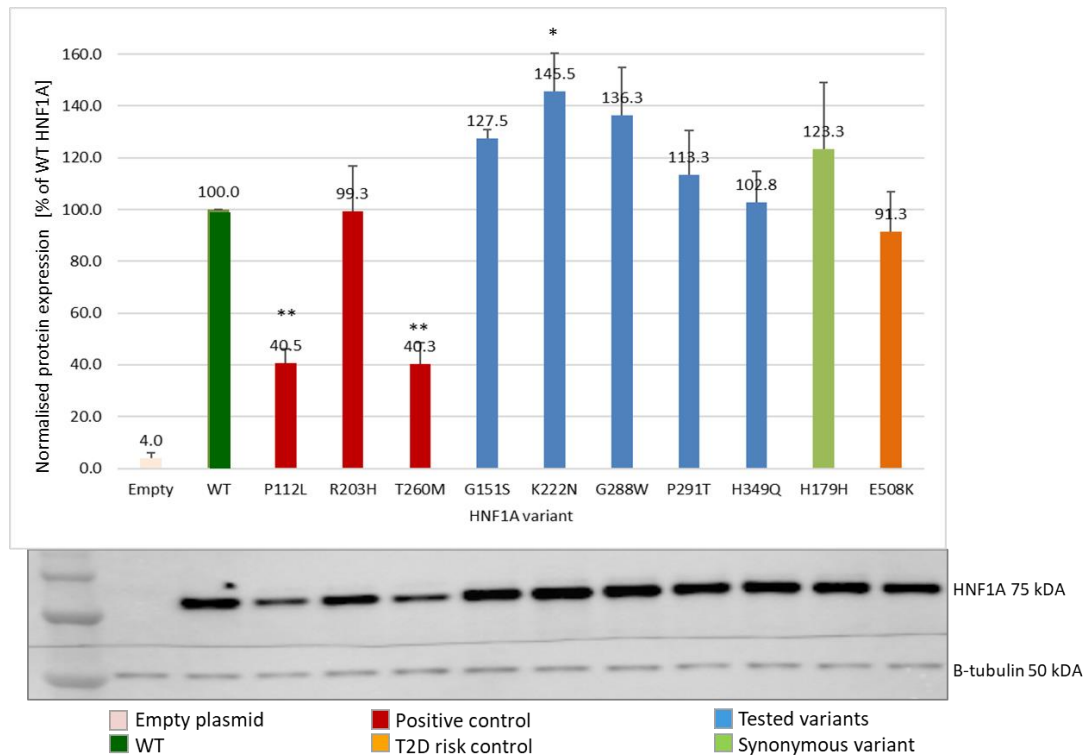


Figure 4.4.2 Protein expression of WT and variants HNF1A using densitometry from western blotting presented as a mean ($\pm$ SEM) percentage of the wild-type (WT) from four biological replicates. B-tubulin was used as a control to assure equal amount of protein lysate has been loaded. One-way ANOVA was used to compare means; $**p=0.001$, $*p<0.01$. At the bottom, an example of a gel image with bands order corresponding to the bars on the chart above. HNF1A bands at the top of the gel and B-tubulin (technical control) at the bottom of the gel with protein ladder on the left.

4.4.4 DNA binding of HNF1A protein variants using Electrophoretic mobility shift assay

The variant p.G151S had significantly reduced DNA binding [$19.2 (\pm 4.0)$ % of the WT, $p=2 \times 10^{-6}$] which was comparable to all three positive controls (p.P112L, p.R203H and p.T260M) (Table 4.4.3, Figure 4.4.8). The mean DNA binding of protein variant p.K222N and p.G288W was stronger than the WT HNF1A [$135.4 (\pm 18.2)$ % of WT, $p=0.016$ and $129.4 (\pm 24.2)$ %, $p=0.042$ respectively].

Variant	n1	n2	n3	n4	mean [% WT]	SD	SEM	p value
Empty	6.7	1.8	13.1	5.0	6.6	4.7	2.4	1.2×10^{-7}
WT	100.0	100.0	100.0	100.0	100.0	0.0	0.0	n/a
P112L	10.2	4.3	11.3	10.7	9.1	3.3	1.6	2.0×10^{-7}
R203H	19.6	19.3	19.2	23.9	20.5	2.3	1.1	2.0×10^{-6}
T260M	13.5	6.3	10.5	9.8	10.0	3.0	1.5	2.6×10^{-7}
G151S	19.2	13.7	21.0	23.0	19.2	4.0	2.0	2.0×10^{-6}
K222N	181.5	138.8	128.0	93.2	135.4	36.4	18.2	0.016
G288W	197.5	125.1	111.0	84.2	129.4	48.5	24.2	0.042
P291T	125.9	90.3	83.2	97.0	99.1	18.7	9.3	0.943
H349Q	73.5	74.7	94.2	60.3	75.7	14.0	7.0	0.088
H179H	79.0	n/a	86.2	64.9	76.7	10.8	6.3	0.106
E508K	n/a	78.1	61.0	56.6	65.2	11.3	6.5	0.037

Table 4.4.3 DNA binding of WT and variant HNF1A using electrophoretic mobility shift assay and Cy5 labelled probe (oligonucleotide from rat albumin promoter). DNA binding expressed as a mean percentage of the WT HNF1A from four biological replicates; SD, standard deviation; SEM, standard error of means; one-way ANOVA used to compare variants; significant $p < 0.05$; n/a, not available.

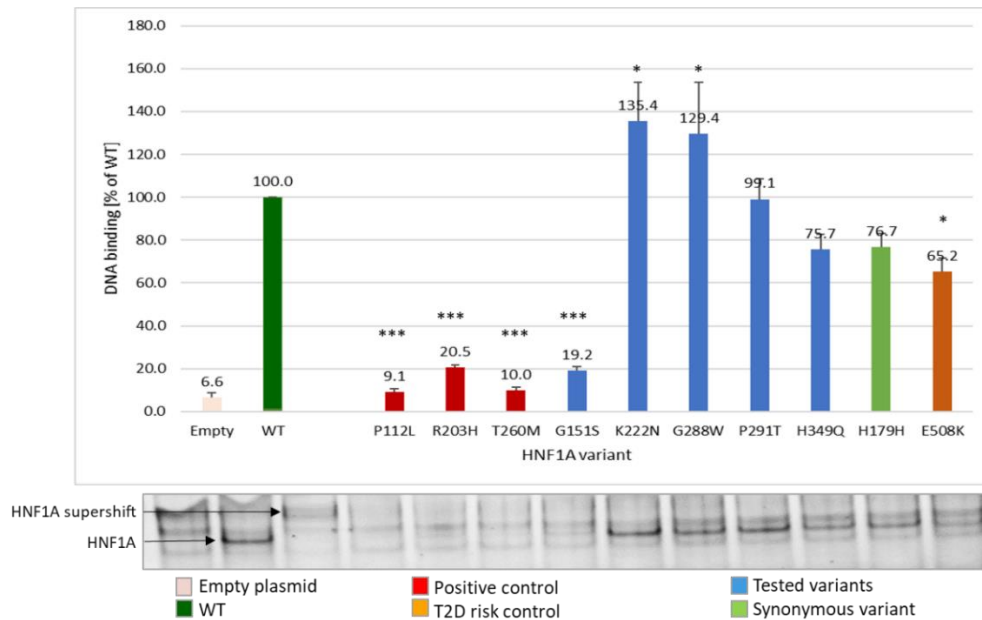


Figure 4.4.3 DNA binding of WT and variant HNF1A using electrophoretic mobility shift assay and Cy5 labelled probe (oligonucleotide from rat albumin promoter). DNA binding expressed as a mean ($\pm$ SEM) percentage of the WT HNF1A from four biological replicates, one-way ANOVA used to compare variants; *** $p < 0.0001$, * $p < 0.05$. At the bottom, an example gel image ($n=1$) with bands

order corresponding to the bars on the chart above showing HNF1A-probe complex (short arrow) and supershift of the bigger complex of HNF1A-probe- α /HNF1A antibody (longer arrow).

The mean DNA binding of variant p.G151S remained similar (15.0% of the WT) when a mean DNA binding from four biological replicates for each variant was normalised for the HNF1A protein amount from the corresponding mean in western blot (Figure 4.4.4). The normalised DNA binding of the remaining four tested variants was not statistically different from the WT HNF1A when protein normalized results were analysed (71.0-101.0 % of WT, $p > 0.05$, Figure 4.4.4).

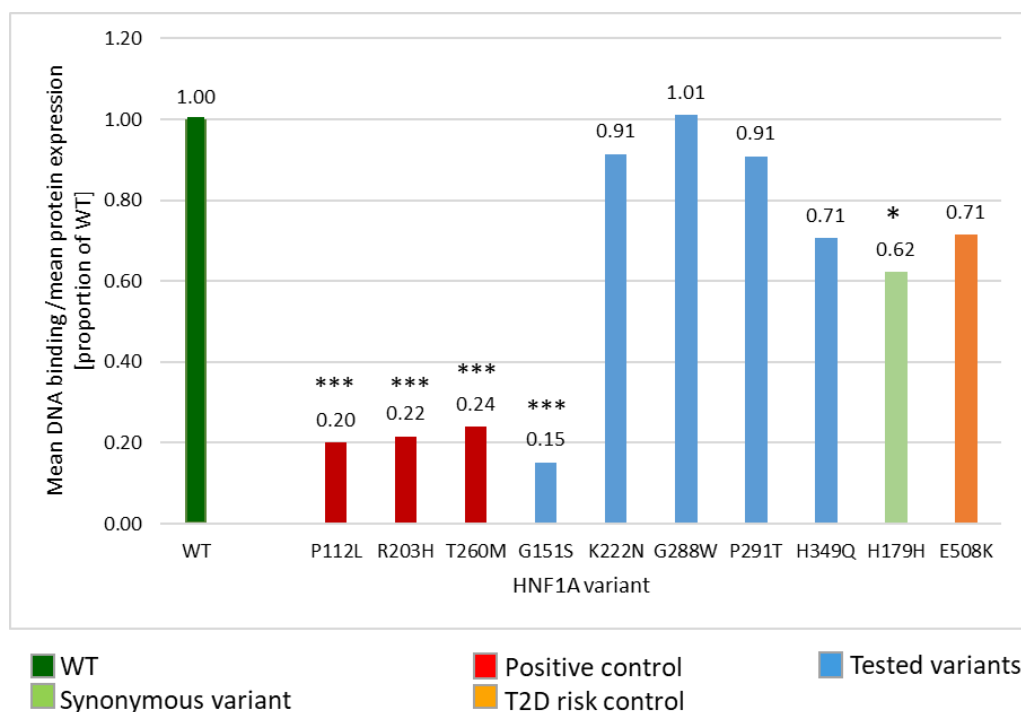


Figure 4.4.4 Mean DNA binding of WT and variant HNF1A from EMSA densitometry normalised for the mean HNF1A protein amount from WB densitometry. *** $p < 0.0001$, * $p < 0.05$.

4.4.5 Summary of the functional studies performed

Functional characterisation of five previously uncharacterised *HNF1A* variant alleles showed that variant p.G151S had significantly reduced DNA binding but HNF1A protein expression and transcriptional activity were similar to the WT. Variants p.G288W and p.P291T had transcriptional activity reduced to 73.3% and 59.2% of the WT HNF1A respectively, but their protein expression and DNA binding were similar to WT HNF1A. Variants p.K222N and p.H349Q both displayed functional results which were comparable to the WT HNF1A.

Table 4.4.4 summarizes features and functional results of ten *HNF1A* alleles included in functional studies (five tested and five control variants).

Protein variant	Variant type	Bioinformatic prediction (SIFT, Polyphen 2 & Provean)	MAF in GnomAD [%]	Co-segregation with diabetes	Functional studies			Current prediction
					TA [% WT]	WB [%WT]	EMSA [%WT]	
Variants tested in this study								
G151S	missense novel	damaging	0	n/a	82.8	127.5	19.2	likely damaging
K222N	missense novel	damaging	0	3 gen of DM*	76.1	145.5	135.4	likely benign
G288W	missense	tolerated	0.007	no	73.3	136.3	129.4	likely benign
P291T	missense	tolerated	0.0008	yes	59.2	113.3	99.1	likely benign
H349Q	missense	tolerated	0.006	n/a	112.5	102.8	75.7	benign
Control variants								
P112L	missense	damaging	0.0004	yes	n/a	40.5	9.1	damaging
H179H	synonymous	tolerated	0.001	n/a	120.5	123.3	76.7	benign
R203H	missense	damaging	0.0008	yes	77.2	99.3	20.5	damaging
T260M	missense	damaging	0	yes	7.8	40.3	10	damaging
E508K	missense	damaging	0.04	yes?	60.9	91.3	65.2	increasing risk of T2D

Table 4.4.4 Features and results of the functional characterisation of ten rare *HNF1A* variants (5 tested and 5 control variants); MAF, minor allele frequency; TA, transcription activity; WB, western blot; EMSA, electrophoretic mobility shift assay; *only *HNF1A* of the proband sequenced; results supporting damaging functional effect of *HNF1A* allele marked in pink.

Work presented in this chapter supports the fact that allele c.451G>A, p.G151S can be causal for HNF1A-MODY phenotype in the individual with diabetes and that alleles p.K222N, G288W, p.P291T and p.H349Q are unlikely to cause the typical MODY phenotype based on the functional results only. However, the clinical phenotype of the individual carrying a particular allele is an outcome of complex genetic and environmental interactions and *in vitro* studies represent only part of the complex human biology.

4.4.6 Updated functional prediction score of *HNF1A* alleles

Including functional characterisation of the *HNF1A* variants (performed in this study and previously published in literature) into the functional prediction score of *HNF1A* alleles, discussed in chapter 3, has changed the prediction score of five alleles tested: p.S3C, p.G47R, p.K222N, p.P291T and p.T515M from likely damaging to likely benign. Table 4.4.5 provides an updated functional prediction score for all alleles found in the current study with functional results included. This changes the score range to 0-4, with a score of 0-1 defining a benign allele, 1.5-2 likely benign, 2.5-3 likely damaging and 3.5-4 damaging variant allele.

Three PTVs which did not have functional characterisation performed (exon 1 del, p.D135fs and p.L377fs) had the final prediction left assigned as damaging. These PTVs are localized in exons 1-6 and likely to cause nonsense-mediated decay, as discussed earlier.

In summary, a total of 9 of 26 alleles (in 13 probands and 2 relatives) are predicted as damaging; 2 of 26 alleles as likely damaging (in 2 probands and 1 relatives), 6 alleles (in 6

proband and 4 relatives) are labelled as likely benign and 9 alleles (in 9 probands and 1 relative) are predicted to be benign (Table 4.4.5, Figure 4.4.5).

Coding DNA variant	Protein variant	Reported as causing MODY with co-segregation	<i>In silico</i> prediction	GnomAD MAF<0.007 % or absence in T2D-GENES	Functional work supportive	Final score	Final prediction
c.-4A>G	n/a	0	0	0	n/a	0	benign
c.1-326del	del exon 1	1	1	1	n/a	3	damaging [#]
c.8C>G	S3C	novel	1	1	n/a	2	likely benign
c.139G>C	G47R	1	0	1	0	2	likely benign
c.142G>A	E48K	1	0	0	inconclusive	1	benign
c.404delA	D135fs	1	1	1	n/a	3	damaging [#]
c.451G>A	G151S	novel	1	1	1	3	likely damaging
c.586A>G	T196A	0	0	0	0	0	benign
c.666G>T	K222N	novel	1	1	0	2	likely benign
c.685C>T	R229*	1	1	1	1	4	damaging
c.751G>A	A251T	0.5	0	1	0	1.5	likely benign
c.779C>T	T260M	1	1	1	1	4	damaging
c.862G>T	G288W	0	0	1	0	1	benign
c.871C>A	P291T	1	0	1	0	2	likely benign
c.872delC	P291fs	1	1	1	1	4	damaging
c.872dupC	G292fs	1	1	1	1	4	damaging
c.1015G>A	G339S	0	0	0	0	0	benign
c.1047C>A	H349Q	0	0	0	0	0	benign
c.1129delC	L377fs	1	1	1	n/a	3	damaging [#]
c.1136_1137delCT	P379fs	1	1	1	1	4	damaging
c.1136C>A	P379H	1	1	1	1	4	damaging
c.1136C>G	P379R	1	1	1	0	3	likely damaging
c.1165T>G	L389V	0	0	0	0	0	benign
c.1544C>T	T515M	0	1	1	0	2	likely benign
c.1748G>A	R583Q	0	0	0	0	0	benign
c.1816G>A	G606S	0	0	0	0	0	benign

Table 4.4.5 Proposed provisional prediction score of the allele phenotypical effect based on the five features discussed above. Final prediction score 0-4 (0-1, benign; 1.5-2, likely benign; 2.5-3, likely damaging; 3.5-4, damaging allele); n/a, not available; [#]PTVs with no functional data available but likely to undergo a nonsense mediated decay labelled as “damaging” instead of “likely damaging” as score would suggest.

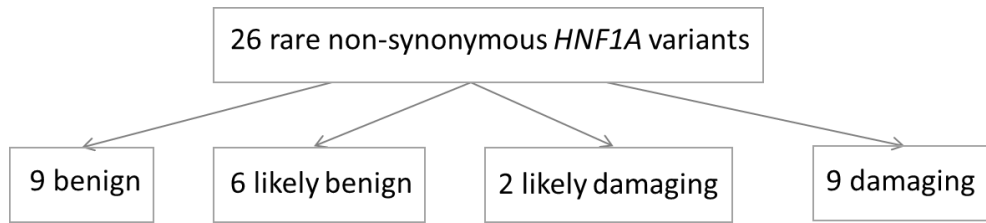


Figure 4.4.5 The final prediction score for 26 rare non-synonymous *HNF1A* variant alleles found in this study.

4.5 Discussion

Functional characterisation described in this chapter showed that one of the five *HNF1A* alleles investigated (p.G151S) had significantly reduced DNA binding (15.0% of the WT) consistent with its localisation in the DNA binding domain and supporting the prediction that this allele can cause the *HNF1A*-MODY phenotype.

However, it was surprising that this variant also showed a non-significant reduction in transactivation potential (82.7% of the WT) when rat albumin promoter and HeLa cell line were used. This result is similar to that of variant p.R203H (TA of 77.2% of the WT and DNA binding 20.5% of the WT in my experimental model), which had previously been reported in the families with MODY phenotypes and autosomal dominant inheritance of diabetes (Awata, Kurihara et al. 1998, Ng, Cockburn et al. 1999). The discrepancy between transactivation results similar to the WT *HNF1A* and significantly affected DNA binding of these protein variants is unclear. Typically, a protein variant with significantly reduced DNA binding would be expected to result in significantly reduced transcriptional activity, taking into account that DNA binding is the first step of transactivation assay. This difference could be an outcome of the experimental model used.

Dr Hassanali from our group investigated transactivation of variant p.R203H using HeLa cells and *HNF4A P2* promoter. In this experimental model, TA of variant p.R203H was significantly reduced when compared to the model using HeLa cells and human albumin promoter (mean TA 20% of the WT vs 60% of the WT HNF1A respectively) (Hassanali, unpublished). The above results suggest that using a β -cell specific promoter such as *HNF4A P2* or *INS* promoter could better reflect human islet biology. Using human instead of rat albumin promoter did not change transactivation significantly (60% vs 77.2% of the WT) as sequence of albumin promoter is highly conserved.

Another explanation for the discrepancy between DNA binding and transactivation results of p.R203H and p.G151S would be a presence of other biological factors in the cell physiology which could rescue potential defects in DNA transcription. Electrophoretic mobility shift assay used in my experiments tested binding of HNF1A variant to the short fragment of rat albumin DNA only. Hence, potential co-factors might have been absent.

The individual carrying allele p.G151S in the current study was diagnosed with diabetes at 35 years of age and was initially only treated with diet. Ten years after diagnosis of diabetes, her BMI was 26.3kg/m², she was treated with metformin, glimepiride (SU) and vildagliptin with HbA1c of 6.3%. The effect of adding sulphonylurea is unknown in this case. Her mother and maternal grandmother have diabetes but were not available for sequencing. Hence, clinical features of this proband are not typical HNF1A-MODY phenotype but borderline for MODY and T2D.

Functional characterisation of *HNF1A* variants from my study and that reported in previous literature reflects the functional spectrum of *HNF1A* alleles and correlates with the spectrum of the clinical phenotype ranging from typical HNF1A-MODY to T2D. Presented in this chapter and previously published results support the fact that alleles such as p.P112L and p.T260M cause the HNF1A-MODY phenotype (Bjorkhaug, Ye et al. 2000).

The family of the proband with allele p.T260M in the current study is an example of how a damaging *HNF1A* allele can cause a typical MODY phenotype. The proband, his sister and their father presented with diabetes at 16, 18 and 24 years of age, all were slim at diagnosis and sensitive to a small dose of gliclazide or glibenclamide. His sister had severe recurrent hypoglycaemia after using 40mg of gliclazide daily (half of a starting dose recommended for an individual with type 2 diabetes), that she refused to take for more than 10 years until she was seen in our clinic, reassured and prescribed 20mg of gliclazide daily instead.

Functional studies of *HNF1A* variants such as p.P291T or p.G288W also addressed the interpretation of intermediate results (60-80% of the WT HNF1A protein) which in fact applies to the majority of rare missense *HNF1A* variants (Najmi, Aukrust et al. 2016). Those results were not statistically significant when compared to the WT HNF1A but could represent the middle of the functional continuum between alleles causing MODY phenotype and those which increase the risk of T2D. This is especially true for *HNF1A* variants with reduced penetrance and localised in exons 8-10 such as p.T515M or

p.R583Q. Individuals carrying variants in exon 8-10 usually present with diabetes later in life than subjects with variants in exons 1-7 and classical MODY phenotype (Harries, Ellard et al. 2006, Bellanne-Chantelot, Carette et al. 2008). Variants such as p.H349Q or p.G606S had been reported as ethnic-population specific and performed in the functional studies similarly to the WT HNF1A. Therefore, these alleles are likely to be neutral to the HNF1A protein function.

Although functional studies using transfected human cell lines could provide some insight into biological processes in humans, they represent only a part of the physiological process tested. Human biological pathways are complex and the results from an experimental model often cannot be extrapolated directly to human disease.

For that reason, showing that the particular allele is functional experimentally and could potentially cause the MODY phenotype raises the question: does this allele cause MODY in a particular individual with diabetes? Due to a spectrum of allelic functional effects, individuals carrying *HNF1A* variants show variation of the clinical phenotype even within the same family (Bellanne-Chantelot, Levy et al. 2011, Fajans and Bell 2011). This results from the influence of other genetic, physiological and environmental factors such as diet or exercise.

An example of a different penetrance *HNF1A* allele is p.P291T from this study and the Oxford MODY clinic. First proband carrying this allele presented with young adult onset diabetes in the 3rd decade of his life, was obese at the time of diagnosis, did not respond to sulphonylurea treatment and had progressed quickly to the insulin treatment. The

mother and cousin of the proband are also obese which is suggestive of the influence of lifestyle and other shared genetic factors. However, another individual carrying the same variant allele and seen in our clinic presented with diabetes at the age of 15 years with normal body weight consistent with a typical MODY phenotype.

The methods used in my functional studies had been widely utilised. The reasonably narrow standard error of means (SEM) of my work results made them reproducible and similar to recently published functional characterisation of *HNF1A* variants (Balamurugan, Bjorkhaug et al. 2016, Najmi, Aukrust et al. 2016). The variation in all assays used in the performed functional studies could be explained by multiple experiment-related factors such as plating density, transfection efficacy, growth of HeLa cells after transfection and accuracy in pipetting at different stages of the experiment. The variation of WB and EMSA results is also explained by differences in protein stability following the freezing-thawing process.

However, new methods for functional characterisation of alleles are being developed and patient-derived induced pluripotent stem cells (iPSC) provide a promising tissue-specific model for human disease. The human iPSC had been derived from skin fibroblasts of subjects with MODY which have been shown to be similar to human embryonic stem cells and could be reprogrammed to provide a β -cell model (Teo et al. 2013).

Another fast-developing genome-editing technology using clustered, regularly interspaced, short palindromic repeat (CRISPR)-Cas9 endonuclease enables the targeted

modification of DNA sequence and provides an efficient tool for a large variety of functional assessments (Shalem et al. 2015).

4.6 Conclusions

Functional characterisation of *HNF1A* variants presented in this chapter reflected the functional spectrum of *HNF1A* alleles which correlates with a spectrum of clinical phenotypes of individuals with C-peptide positive and β -cell antibody negative diabetes. The final phenotype of the individual is dependent on interactions between genetic and environmental factors and can vary even between subjects with the same *HNF1A* variant allele.

Chapter 5

Biomarkers of HNF1A-MODY and translational follow-up

The contents of this chapter have been presented at the EASD-Study Group on Genetics of Diabetes meeting and the 53rd Annual Meeting of the European Association for the Study of Diabetes:

- *Juszczak A, Pavić T, Bennett AJ et al. High-sensitivity CRP and antennary fucosylated glycans as biomarkers of HNF1A-MODY. SGGD 2017, Leiden, Netherlands (an oral presentation).*
- *Gornik O, Pavić T, Juszczak A, Pape-Medvidović E, Bennett AJ, Goyn AL, et al. Plasma N-glycans and C-reactive protein as biomarkers of HNF1A-MODY in young adults. Diabetologia (2017) 60 (Suppl 1):S101.*

A manuscript including context of this chapter has been submitted to the peer reviewed journal and awaiting response.

5.1 Introduction

5.1.1 Previously reported biomarkers of HNF1A-MODY

Chapter 1.5 described the biomarker candidates of HNF1A-MODY reported in previous literature. Of those discussed, high-sensitivity CRP and N-glycan profile showed clinical potential as biomarkers of HNF1A-MODY. In chapter 3, I systematically assessed 26 rare non-synonymous *HNF1A* variants found by sequencing 1032 subjects with young adult-onset non-autoimmune diabetes from the UK and Croatia. In chapter 4, I discussed the functional characterisation of *HNF1A* alleles identified in my study, referencing previous literature as well as my own findings. That work complemented assigning the likely phenotypical effect of these *HNF1A* alleles and allowed me to divide the subjects in the study into groups depending on that effect, discussed in section 4.4.7 of chapter 4.

This chapter describes the assessment of C-reactive protein and 3-antennary fucosylated N-glycans as biomarkers of HNF1A-MODY in those groups determined by the *HNF1A* allele status.

5.1.1.1 C-reactive protein

The study by Thanabalasingham *et al.* described in detail in chapter 1, section 1.6.2.1 showed that CRP measured by high-sensitivity assays (hsCRP) in individuals from 7 European centres, was significantly lower in subjects with HNF1A-MODY than in those with type 2 diabetes [median CRP in HNF1A-MODY in all countries 0.03-0.25 (0.03-0.98) mg/L vs 1.44-5.21 (0.40-16.3) mg/L in T2D, $p < 1.3 \times 10^{-5}$] (Thanabalasingham, Shah *et al.*

2011). CRP discriminated subjects with HNF1A-MODY from those with T2D well with the ROC curve-derived C-statistics of 0.79-0.97 for the hsCRP cut-off 0.25-0.5mg/L depending on three different assays used.

Bellanne-Chantelot *et al.* reported that hsCRP did not help in differentiating participants with HNF1A-MODY (n=143) from slim individuals with familial young-onset T2D (n=310) (Bellanne-Chantelot et al. 2016). However, the median hsCRP was significantly lower in individuals with HNF1A-MODY when compared to slim subjects with young-onset T2D (0.65mg/L vs 1.42mg/L, $p<0.0001$). The authors also concluded that there is no need for using hsCRP in differentiating HNF1A-MODY from typical (overweight, median BMI 29.9kg/m²) T2D individuals (n=215), as clinical features on their own achieved C-statistic of 0.99. This is consistent with our clinical practice when patients with typical features of T2D would not be considered for genetic testing.

5.1.1.2 Plasma N-glycan profile

HNF1A was reported as a key regulator of plasma protein fucosylation and its involvement in both guanosine-diphosphate-fucose synthesis and in the activity of antennary fucosyltransferase 3, 5 and 6 (Lauc, Essafi et al. 2010) was described in chapter 1, section 1.5.2.

The study by Thanabalasingham *et al.* described in detail in chapter 1, section 1.6.2.2 showed that subjects with HNF1A-MODY had marked alterations of plasma N-glycan profile measured by high-performance liquid chromatography method, which separated

16 glycan groups and derived 13 desialylated glycans (DG) (Thanabalasingham, Huffman et al. 2013). The desialylated glycan DG9 index, which summarized the proportion of 3-antennary glycans that were fucosylated [DG9/(DG8+DG9)], had been shown to be significantly lower in individuals with HNF1A-MODY when compared to those with typical T2D [0.09 (0.06-0.13) vs 0.25 (0.18-0.33)] and type 1 diabetes [0.09 (0.06-0.13) vs 0.28 (0.20-0.34), $p \leq 1 \times 10^{-34}$ for both comparisons]. The C-statistics were 0.94 and 0.91 respectively to distinguish HNF1A-MODY from T1D and T2D. The DG9 index cut-off <0.16% had a sensitivity of 88% and specificity of 81% differentiating subjects with HNF1A-MODY from those with typical T2D.

However, the studies described above used groups of well-defined individuals with an extreme phenotype who had already been assigned a genetic diagnosis, which is hardly the case in a real-life clinical setting. Therefore, I assessed a discriminative performance of CRP and N-glycans in the subjects resembling those from young adult diabetes clinic.

5.1.3 Aims of the study

In the work described in this chapter, I aimed to:

1. Assess the value of hsCRP and N-glycan profile as diagnostic markers of HNF1A-MODY by testing their performance in an unselected population with young adult-onset non-autoimmune diabetes;
2. Examine if either of the two examined biomarkers could be used as predictor of the functional effect of the *HNF1A* variant allele;
3. Assess a translational potential of hsCRP and the best performing N-glycan.

5.2 Study subjects and methods

5.2.1 Study participants and inclusion criteria

Subjects with C-peptide positive, GAD antibody negative diabetes participating in this study were recruited from the UK and Croatia, as described in chapter 2 section 2.1.1 and further in chapter 3, section 3.2.1. The total number of recruited individuals was 1058 (581 from the UK and 477 from Croatia). The clinical characteristics of these subjects divided according to the country of recruitment are given in section 2.1.1 and divided according to the *HNF1A* allele functional effect in Table 5.2.1 below.

All participants had *HNF1A* sequenced and hsCRP and *N*-glycan profile measured. Twenty six subjects were excluded due to failed sequencing, leaving a total number of participants of 1032. For the analysis of hsCRP, three subjects with missing CRP measurements and 116 subjects with CRP >10mg/L were excluded as this was likely to be due to co-existing infection. For the analysis of glycans, 54 participants were excluded due to missing glycan data and one due to end-stage renal failure, leaving 913 subjects for CRP and 977 subjects for glycan investigations.

5.2.2 Subject characteristics

Initially, subjects were divided into five groups depending on *HNF1A* allele functional status: four groups with rare non-synonymous *HNF1A* variants (damaging, likely damaging, likely benign and benign alleles) and one without *HNF1A* variants. However, for the purpose of biomarker assessment three groups were analysed: those with

damaging/likely damaging alleles, likely benign/benign alleles and those without *HNF1A* variants. Table 5.2.1 summarises their characteristics.

Subjects with damaging/likely damaging *HNF1A* alleles were younger at diagnosis of diabetes [median 24.5 (18.0-29.5) vs 37.0 (32.0-40.0) years, $p=3.0 \times 10^{-6}$], slimmer [median BMI 25.7 (21.6-29.8) vs 30.4 (26.9-35.2) kg/m², $p=0.001$], had lower C-peptide [median 0.32 (0.27-0.62) vs 0.70 (0.43-0.99) nmol/L, $p=7.6 \times 10^{-5}$], higher HDL-cholesterol [median 1.30 (1.10-1.58) vs 1.15 (0.96-1.35) mmol/L, $p=0.03$] and lower triglycerides [median 1.06 (0.90-1.31) vs 1.60 (1.11-2.38) mmol/L, $p=0.004$] than subjects with no *HNF1A* variants.

Subject group depending on <i>HNF1A</i> variant allele	Damaging/likely damaging allele	Likely benign/benign allele	No <i>HNF1A</i> variant	P value (damaging/likely damaging <i>HNF1A</i> alleles vs all others)	P value (damaging/likely damaging alleles vs no variant)
Number of patients	18	20	994		
Gender (number of men/%)	7 (38.9)	9 (45.0)	609 (61.3)	0.23	0.05
Age at recruitment [median(IQR) years]	35.0 (28.5-53.5)	48.0 (40.5-62.5)	47.0 (41.0-54.0)	0.05	0.02
Age at diagnosis [median(IQR) years]	24.5 (18.0-29.5)	35.0 (26.8-38.3)	37.0 (32.0-40.0)	9.0×10^{-6}	3.0×10^{-6}
BMI [median (IQR) kg/m ²]	25.7 (21.6-29.8)	32.0 (25.7-32.1)	30.4 (26.9-35.2)	0.003	0.001
FPG [median (IQR) mmol/L]	7.0 (5.4-8.5)	7.6 (5.6-9.8)	8.1 (6.5-10.4)	0.14	0.13
HbA1c [median (IQR) %]	7.3 (6.4-7.8)	8.6 (6.4-9.6)	7.5 (6.6-8.8)	0.35	0.31
Duration of DM [median (IQR) years]	10.0 (6.3-22.3)	9.0 (5.0-22.0)	10.0 (5.0-17.0)	0.65	0.69
C-peptide [median (IQR) nmol/L]	0.32 (0.27-0.62)	0.61 (0.29-0.89)	0.70 (0.43-0.99)	2.2×10^{-4}	7.6×10^{-5}
Total cholesterol [median (IQR) mmol/L]	4.8 (3.9-5.2)	4.6 (3.6-5.4)	4.6 (3.9-5.4)	0.98	0.97
HDL [median (IQR) mmol/L]	1.30 (1.10-1.58)	1.10 (0.90-1.50)	1.15 (0.96-1.35)	0.09	0.03
TG [median (IQR) mmol/L]	1.06 (0.90-1.31)	1.43 (0.93-2.28)	1.60 (1.11-2.38)	0.01	0.004
Treatment					
Insulin +/- OHA [number of case (%)]	8 (44.4)	13 (65.0)	408 (41.0)	0.001	0.77
OHA [number of case (%)]	6 (33.3)	2 (10.0)	497 (50.0)		0.16
diet [number of case (%)]	4 (22.2)	5 (25.0)	89 (9.0)		0.05

Table 5.2.1 Subject characteristics (n=1032, all subjects with *HNF1A* sequenced). Continuous variables are given as median (IQR) as non-parametrically distributed, while categorical variables are given as percentages. BMI, body mass index; OHA, oral hypoglycaemic agents; T2D, type 2 diabetes. Significant p value in bold (<0.05), Kruskal-Wallis test used to compare all groups, Mann-Whitney test to compare variables in patients with damaging/likely damaging *HNF1A* alleles versus those with no variant.

Subjects with damaging/likely damaging *HNF1A* variants were treated with insulin with or without OHA as often as participants with no *HNF1A* variants (44.4% vs 41.0%, $p=0.77$), as often with OHA only (33.3% vs 50.0%, $p=0.16$) and more often with diet alone (22.2% vs 9.0%, $p=0.05$) in comparison to the individuals without *HNF1A* variants. In the group with damaging/likely damaging *HNF1A* variants, 11 subjects (64.7%) were treated with sulphonylurea or glinides as mono- or combined therapy, while this treatment applied to 461 individuals (46.4%) in the group without *HNF1A* variants ($p=0.12$).

5.2.3 Sample size calculation for the assessment of biomarkers

I used the Buderer method (Buderer 1996) for calculating the sample size required for a diagnostic test and detailed calculations are provided below. This takes into account the sensitivity and specificity of the test and the precision (0.05 for the 95% confidence intervals (CI) for the sensitivity and specificity, 0.1 for 90% CI). A sample size was calculated for a power of 80%, α of 0.05 and a prevalence of disease (*HNF1A*-MODY) of 5% (Thanabalasingham, Pal et al. 2012). The sensitivity and specificity from previously reported studies were applied (Thanabalasingham, Shah et al. 2011, Thanabalasingham, Huffman et al. 2013), as follows.

5.2.3.1 Calculation of the sample size using sensitivity

$$N = \frac{(TP + FN)}{P}$$

$$TP + FN = Z^2 \times \frac{SN \times (1 - SN)}{W^2}$$

Where N is the sample size; TP , true positive; FN , false negative; Z , stable for confidence interval distribution value of 95% ($Z=1.96$); SN , sensitivity; W , precision; P , prevalence of the disease.

DG9 glycan index with sensitivity 88% and precision 0.1 required a sample size of 811, while hsCRP with sensitivity of 80% and precision of 0.1 required a sample size of 1228.

5.2.3.2 Sample size calculated for specificity

$$N = \frac{FP + TN}{1 - P}$$

$$FP + TN = Z^2 \times \frac{SP \times (1 - SP)}{W^2}$$

The sample size required for specificity of 81% and precision of 0.05 using the DG9 index was 249, while for specificity of 80% and precision 0.05 for CRP it was 259.

Taking into account the above calculations for sensitivity and specificity, I aimed for a sample size of around 1200, being aware that for the glycans study 811 would be sufficient.

5.2.4 DNA extraction and sequencing

I extracted DNA, and amplified and sequenced *HNF1A* as described in chapter 2, section 2.3-2.6. I used Mutation Surveyor version 5.0.1 (Soft Genetics, Cambridge, UK) for the detection of mutations comparing to the reference sequence (NM_000545.5).

5.2.5 Assigning disease causality to *HNF1A* alleles

A systematic assessment of rare *HNF1A* variant alleles was described in detail in chapter 3. In summary, it included clinical features of patients, co-segregation of the variant with diabetes in family members and *in silico* analysis of missense variants using prediction software (SIFT, Polyphen 2 and Provean). The presence of rare variant alleles was searched for in the publically available database of 123,136 exomes and 15,496 whole-genomes in the Genome Aggregation Database (<http://gnomad.broadinstitute.org/>) (Lek, Karczewski et al. 2016) and in the group of non-diabetic controls from the T2D-GENES project (Fuchsberger, Flannick et al. 2016), which also contributed to the ExAC/GnomAD database. The results of the functional work available in the literature were reviewed and functional studies of five previously uncharacterised *HNF1A* variants were performed and presented in chapter 4.

All analysed evidence to assign a functional effect of *HNF1A* alleles was included in the functional prediction score of the allele and gave basis to divide subjects in the study into groups for the assessment of biomarkers as discussed in chapter 4, section 4.4.7 and Table 4.4.2.

5.2.6 C-peptide, GAD antibodies and CRP assays

Measurement of C-peptide, GAD antibodies and CRP using high-sensitivity assay were described in chapter 2, section 2.10. C-peptide and hsCRP were measured in the Biochemistry Laboratory of OUH NHS Trust by Neha Shah. The GAD antibodies were

measured in the Immunology Laboratories in the University of Bristol by Kyla Chandler for patients recruited from the UK and by the Immunology Laboratory staff in the Merkur Hospital in Zagreb for participants recruited from Croatia.

5.2.7 Analysis of plasma N-glycans

Detailed descriptions of plasma sample preparation, N-glycan labelling and N-glycome separation using the ultra-performance liquid chromatography method are described in chapter 2, section 2.11. N-glycan profile was measured in the Laboratories of the Faculty of Pharmacy and Biochemistry, University of Zagreb by Ms Tamara Pavic.

Briefly, plasma samples were denatured, glycans released using endoglycosidase, and the released N-glycans were labelled with fluorescent 2-aminobenzamine. After clean-up, N-glycans fixed on the GHP plate membrane were eluted and samples injected consecutively to the silica-based column of the Waters Acquity UPLC instrument for the chromatographic separation based on the mass and charge of individual glycans. The HILIC-UPLC method separated 42 different glycan peaks (Figure 5.1) and the separation was improved compared to the HPLC method used previously (separated 16 glycan peaks) (Thanabalasingham, Huffman et al. 2013). The area under the curve (AUC) of each N-glycan peak was expressed as a percentage of total N-glycans integrated AUC. There was no need for measuring desialylated glycans using the HILIC-UPLC method as individual untreated N-glycans performed well in differentiating my study groups.

5.2.8 Statistical analysis

I used SPSS version 23 for the analysis of the data. Continuous data with normal distribution (e.g. current age) were presented as means (SD), those with non-parametric distribution as medians (IQR). The t-test was used to compare means, Kruskal-Wallis test was applied to compare median variables from subjects with *HNF1A* variants versus all other groups together, while the Mann-Whitney test was used to compare any two groups. Differences of frequencies for categorical variables were tested using the Chi-squared or Fisher's exact test. Cases with missing values were excluded pairwise. A p value of <0.05 was considered statistically significant with the exception of the analysis of 42 different N-glycans, where the p value was corrected for false discovery rate of 10% using the Benjamini-Hochberg procedure. All $p < 0.05$ were ranked from the smallest to the highest and adjusted p value calculated using the formula:

$$p_{BH} = \frac{i}{m} \times Q$$

Where i denotes rank, m the number of different tests (hypotheses) investigated and Q the false discovery rate. This reduced the significant p value to <0.01.

Screening test statistics were calculated as follows from a 2 x 2 table including true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN) for both biomarker candidates:

$$SN = TP/(TP + FN)$$

$$SP = TN/(TN + FP)$$

$$PPV = TP/(TP + FP)$$

$$NPV = TN/(TN + FN)$$

$$PLR = SN/(1 - SP)$$

$$NLR = (1 - SN)/SP$$

Where *SN* is sensitivity; *SP*, specificity; *PPV*, positive predictive value; *NPV*, negative predictive value; *PLR*, positive likelihood ratio; *NLR*, negative likelihood ratio.

5.3 Results

5.3.1 High-sensitivity CRP in subjects with rare *HNF1A* variants

The performance of the biomarkers as potential screening tests was analysed first in all subjects with rare non-synonymous *HNF1A* variants versus those without *HNF1A* variants. Then, taking into account the systematic assessment of *HNF1A* alleles, the comparison was performed between individuals who had damaging/likely damaging *HNF1A* alleles versus those without *HNF1A* variants.

Three subjects with benign/likely benign *HNF1A* alleles and 116 without rare *HNF1A* variants had CRP above 10mg/L, and therefore were excluded as CRP was likely to be elevated due to associated infection/inflammation. Median hsCRP of individuals with rare *HNF1A* variants (n=35), regardless of assigned functional effect, was lower than in those without *HNF1A* variants (n=878) [0.80 (0.20-2.00) mg/L versus 1.71 (0.60-3.90) mg/L, Mann-Whitney p=0.001, Figure 5.3.1 A].

The difference became even more significant when participants with damaging/likely damaging *HNF1A* alleles (n=18) were compared to those without *HNF1A* variants (n=878) [median CRP 0.21 (0.04-0.75) mg/L versus 1.70 (0.60-3.90) mg/L, Mann-Whitney p= 1.6×10^{-5} , Figure 5.3.1 B]. The results did not change when only probands were included [median CRP 0.21 (0.06-0.70) mg/L in subjects with damaging/likely damaging *HNF1A* alleles vs 1.70 (0.80-3.70) mg/L in those without *HNF1A* variants, p= 0.80×10^{-5}].

Median CRP was lower in subjects with protein truncating *HNF1A* variants (PTVs, n=12) than in those with missense *HNF1A* variants (n=23) [0.16 (0.07-0.63)mg/L versus 1.25 (0.30-2.73) mg/L, p=0.01]. This difference became non-significant when only damaging/likely damaging variants were analysed [0.16 (0.07-0.63) mg/L versus 0.70 (0.01-1.75) mg/L, p=0.62].

Only three *HNF1A* variants were located in exons 7-10, therefore affecting only isoform A of the HNF1A protein. Thus, comparison of CRP in subjects with *HNF1A* variants affecting all isoforms of the HNF1A protein versus those affecting only isoform A was not reliable.

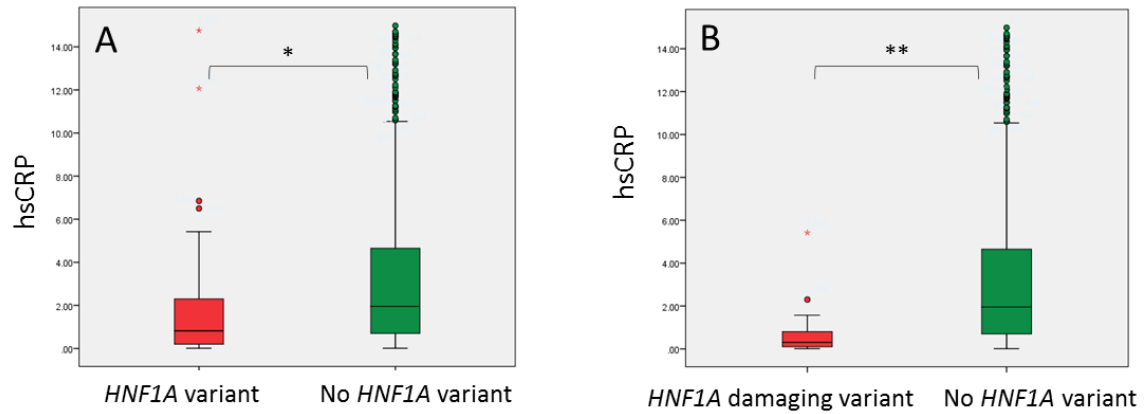


Figure 5.3.1 Median hsCRP in subjects with rare *HNF1A* variant (A, *n*=35) and in subjects with damaging/likely damaging *HNF1A* alleles (B, *n*=18) both versus subjects without *HNF1A* variant (*n*=878); * *p*<0.01, ** *p*<0.001.

5.3.2 Plasma N-glycan profile in subjects with rare *HNF1A* variants

I analysed N-glycans in a total of 977 participants, including 35 subjects with *HNF1A* variants (1 person with end-stage renal failure and damaging *HNF1A* allele and 2 with benign alleles and missing glycan data, were excluded). Thirty eight individual N-glycan values were derived from the area under the curve (AUC) of 42 chromatographic peaks (four N-glycan values represented two adjacent peaks) and presented as a percentage of total N-glycome AUC from each sample chromatogram.

Five of 38 glycan measures (median GP25, 30, 36, 38 and 42) were significantly lower in the subjects with rare *HNF1A* variants (*n*=35) than in participants without the *HNF1A* variants (*n*=942) (Benjamini-Hochberg adjusted *p*<0.01, Table 5.3.2). Out of those 5 glycans, GP30, GP36 and GP38 showed the most significant difference:

- GP30 0.60 (0.41-0.99) vs 0.99 (0.75-1.27) % (Mann-Whitney $p=1.6 \times 10^{-5}$),
- GP36 1.44 (1.08-2.25) vs 2.16 (1.65-2.71) % ($p=2.7 \times 10^{-4}$),

- GP38 0.22 (0.17-0.34) vs 0.32 (0.25-0.40) % ($p=1 \times 10^{-4}$, Figure 5.3.2 top panel).

The results remained strongly significant when relatives of subjects with *HNF1A* variants were excluded ($p=3.0 \times 10^{-7}$, $p=3.4 \times 10^{-5}$ and $p=1.1 \times 10^{-5}$ respectively). Each peak significantly lower in the participants with rare *HNF1A* variants contained 3- or 4- antennary fucosylated N-glycan (Figure 5.2.1). GP35 is also 3-antennary fucosylated but it was not significantly different between the groups ($p=0.21$).

Median GP30 and GP38 were lower in subjects with protein truncating *HNF1A* variants (PTVs, $n=12$) than in those with missense *HNF1A* variants ($n=23$):

- GP30 0.48 (0.37-0.56) % versus 0.79 (0.33-0.70) %, $p=0.009$,
- GP38 0.17 (0.14-0.23) % versus 0.27 (0.16-0.25) %, $p=0.002$.

There was no difference in levels of GP36 in both groups. N-glycan measures which were significantly higher in subjects with *HNF1A* variants when compared to those without *HNF1A* variants were GP27.28, GP33 and GP40 with median:

- GP27.28 of 2.51 (2.27-2.95) vs 2.26 (1.90-2.60) % (Mann-Whitney $p=0.009$),
- GP33 of 4.88 (4.31-5.25) vs 4.24 (3.65-4.94) % ($p=0.006$),
- GP40 of 0.37 (0.29-0.43) vs 0.31 (0.25-0.36) % (0.001).

All these glycan peaks were lacking antennary fucose.

N-glycan	Damaging/likely damaging <i>HNF1A</i> alleles	Benign/likely benign <i>HNF1A</i> alleles	All rare (MAF<1%) <i>HNF1A</i> variants	No rare <i>HNF1A</i> variants
GP1	3.94 (3.31-6.50)	5.38 (3.70-6.47)	4.43 (3.59-6.50)	4.64 (3.62-6.03)
GP2	1.83 (1.55-2.30)	2.11 (1.79-2.45)	2.04 (1.71-2.34)	2.04 (1.74-2.43)
GP3	0.07 (0.06-0.09)	0.08 (0.06-0.11)	0.07 (0.06-0.09)	0.08 (0.06-0.10)
GP4	4.42 (4.11-5.46)*	4.04 (2.89-4.40)	4.23 (3.35-4.79)	3.85 (3.21-4.49)
GP5	2.59 (2.06-2.81)*	2.08 (1.50-2.55)	2.35 (1.72-2.67)	2.09 (1.63-2.51)
GP6	1.34 (1.13-1.63)	1.32 (1.11-1.45)	1.33 (1.15-1.60)	1.31 (1.12-1.54)
GP7	1.09 (0.99-1.12)	1.08 (0.98-1.20)	1.08 (0.98-1.16)	1.07 (0.99-1.17)
GP8	2.00 (1.87-2.11)	1.95 (1.80-2.16)	1.98 (1.86-2.15)	2.02 (1.81-2.23)
GP9	0.12 (0.11-0.14)	0.12 (0.11-0.14)	0.12 (0.11-0.14)	0.12 (0.10-0.13)
GP10.11	5.25 (3.82-6.07)*	3.76 (3.49-4.51)	4.40 (3.51-5.40)	4.00 (3.35-4.82)
GP12	1.14 (1.00-1.20)	0.98 (0.90-1.29)	1.05 (0.91-1.20)	0.99 (0.84-1.14)
GP13	1.21 (1.13-1.39)	1.16 (1.08-1.38)	1.21 (1.09-1.38)	1.21 (1.09-1.33)
GP14	0.83 (0.69-0.99)	0.69 (0.61-0.87)	0.79 (0.64-0.94)	0.71 (0.59-0.85)
GP15	14.15 (13.35-14.91)	14.03 (13.22-15.50)	14.04 (13.24-15.11)	14.41 (13.50-15.46)
GP16	0.41 (0.36-0.45)	0.40 (0.40-0.47)	0.41 (0.37-0.46)	0.40 (0.36-0.44)
GP17.18	6.26 (5.24-6.65)	6.16 (5.59-6.50)	6.16 (5.44-6.59)	5.86 (5.26-6.46)
GP19	2.15 (1.80-2.27)	1.99 (1.79-2.56)	2.02 (1.80-2.37)	1.89 (1.58-2.23)
GP20.21	3.21 (2.78-3.51)*	2.69 (2.41-3.17)	3.04 (2.65-3.39)	2.83 (2.51-3.12)
GP22	1.07 (1.00-1.16)	1.14 (1.10-1.22)	1.11 (1.04-1.19)	1.15 (1.06-1.25)
GP23	22.88 (22.54-24.05)	23.72 (22.23-24.92)	22.97 (22.35-24.84)	23.86 (22.25-25.53)
GP24	0.53 (0.48-0.58)	0.56 (0.49-0.64)	0.55 (0.48-0.63)	0.58 (0.53-0.64)
GP25	3.37 (2.79-3.82)**	4.09 (3.77-4.75)	3.77 (3.14-4.21)*	4.16 (3.59-4.79)
GP26	1.89 (1.44-2.07)	2.01 (1.83-2.21)	1.93 (1.66-2.15)	1.84 (1.50-2.22)
GP27.28	2.86 (2.46-3.07)**	2.33 (1.60-2.57)	2.51 (2.27-2.95)	2.26 (1.90-2.60)
GP29	2.32 (2.13-2.54)	2.02 (1.84-2.57)	2.20 (1.95-2.48)	2.19 (1.89-2.56)
GP30	0.43 (0.34-0.57)***	1.00 (0.64-1.29)	0.60 (0.41-0.99)**	0.99 (0.75-1.27)
GP31	0.77 (0.68-0.83)**	0.69 (0.52-0.75)	0.75 (0.65-0.80)	0.65 (0.55-0.75)
GP32	0.16 (0.14-0.19)	0.15 (0.12-0.20)	0.16 (0.13-0.20)	0.15 (0.13-0.17)
GP33	5.21 (4.62-5.62)**	4.35 (3.49-4.88)	4.88 (4.31-5.25)*	4.24 (3.65-4.94)
GP34	0.43 (0.39-0.48)	0.48 (0.37-0.54)	0.44 (0.39-0.50)	0.46 (0.36-0.57)
GP35	1.15 (1.03-1.45)	1.42 (0.99-1.57)	1.20 (0.99-1.53)	1.31 (1.09-1.56)
GP36	1.17 (1.01-1.52)***	2.23 (1.43-3.25)	1.44 (1.08-2.25)**	2.16 (1.65-2.71)
GP37	0.36 (0.31-0.39)	0.37 (0.32-0.41)	0.36 (0.32-0.41)	0.37 (0.32-0.43)
GP38	0.17 (0.15-0.23)***	0.32 (0.23-0.46)	0.22 (0.17-0.34)**	0.32 (0.25-0.40)
GP39	0.51 (0.40-0.58)	0.48 (0.43-0.53)	0.49 (0.41-0.55)	0.50 (0.43-0.53)
GP40	0.42 (0.32-0.47)**	0.33 (0.24-0.38)	0.37 (0.29-0.43)*	0.31 (0.25-0.36)
GP41	0.69 (0.63-0.79)	0.68 (0.60-0.79)	0.68 (0.62-0.79)	0.68 (0.60-0.76)
GP42	0.38 (0.31-0.42)**	0.47 (0.39-0.64)	0.40 (0.34-0.52)*	0.50 (0.42-0.61)

Table 5.3.2 All plasma median (IQR) N-glycans in subjects divided into 4 groups depending on *HNF1A* variants: damaging/likely damaging (n=17), benign/likely benign (n=18), all *HNF1A* variants (first two groups added, n=35) and no *HNF1A* variants (n=942). Mann-Whitney test to compare N-glycans in subjects with damaging/likely damaging alleles and all *HNF1A* variants versus N-glycans in subjects without *HNF1A* variants; * $p<0.01$, ** $p<0.001$, *** $p<1\times10^{-7}$. Significant $p<0.01$ after Benjamini-Hochberg procedure applied.

Subsequently, individuals with damaging/likely damaging *HNF1A* alleles (probable HNF1A-MODY, n=17, one subject with end-stage renal failure excluded) were compared to subjects without the *HNF1A* variants (n=942). In this analysis, levels of glycans in individuals with damaging/likely damaging *HNF1A* alleles differed significantly in 13 of 38 N-glycan measures (GP4, 5, 10.11, 20.21, 25, 27.28, 30, 31, 33, 36, 38, 40 and 42) but the same five glycan peaks (all 3-antennary fucosylated) were significantly lower in individuals with damaging/likely damaging *HNF1A* alleles versus subjects without the *HNF1A* variants. The remaining 7/13 significantly different N-glycan peaks were higher in those with damaging/likely damaging *HNF1A* alleles than in those without *HNF1A* variants and were not 3-antennary fucosylated (Table 5.3.2).

GP30, GP36 and GP38 had again the most significant p values in comparing the groups with damaging/likely damaging *HNF1A* alleles and those without *HNF1A* variants with median:

- GP30 of 0.43 (0.34-0.57) vs 0.99 (0.75-1.27) % (Mann-Whitney $p=4.2 \times 10^{-9}$),
- GP36 of 1.17 (1.01-1.52) vs 2.16 (1.65-2.71) % ($p=8.1 \times 10^{-8}$),
- GP38 of 0.17 (0.15-0.23) vs 0.32 (0.25-0.40) % ($p=9.3 \times 10^{-9}$, the bottom panel of Figure 5.3.2).

The results remained similar when relatives of patients with *HNF1A* variants were excluded ($p=5.2 \times 10^{-9}$, $p=8.5 \times 10^{-8}$ and $p=9.6 \times 10^{-9}$ respectively).

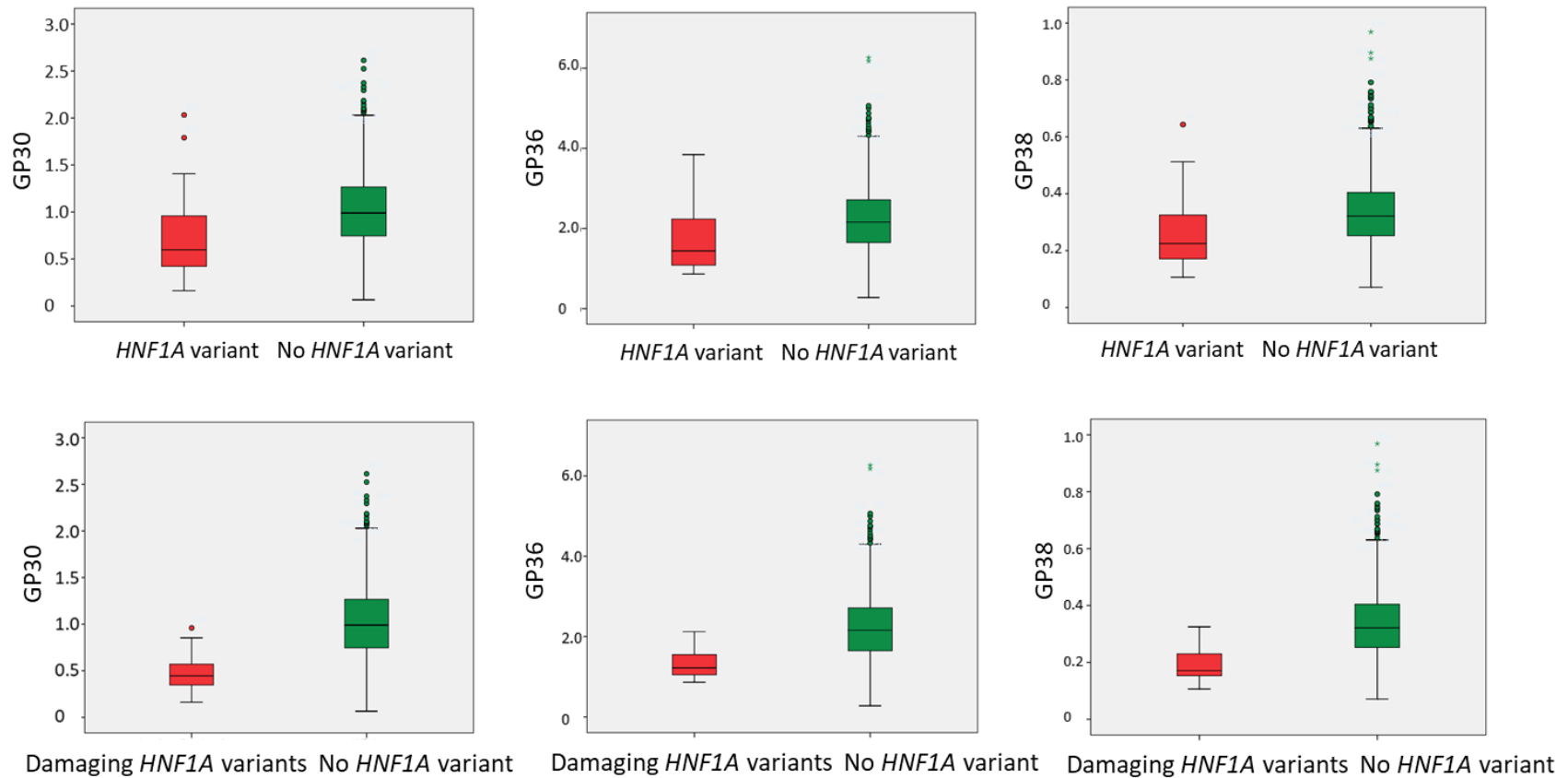


Figure 5.3.2 Median plasma N-glycans: GP30, GP36 and GP38 [%] in subjects with HNF1A variant (n=35) versus no HNF1A variant (n=942, top panel), in subjects with damaging/likely damaging HNF1A variant (n=17) versus no HNF1A variant (n=942, bottom panel). All p values $\leq 2.7 \times 10^{-4}$.

There was no difference in levels of GP30, GP36 and GP38 when subjects with protein truncating *HNF1A* variants were compared to likely damaging missense *HNF1A* variants (p=0.55-0.68).

5.3.3 Discriminating HNF1A-MODY from young-adult onset type 2 diabetes using plasma glycans and hsCRP

High-sensitivity CRP, GP30, 36 and 38 were significantly lower in subjects with damaging/likely damaging *HNF1A* alleles when compared to the participants without the *HNF1A* variants. Hence, they underwent receiver operating characteristic (ROC) curve analysis to assess how they select subjects with HNF1A-MODY from those with young-onset T2D (Figure 5.3.3).

The glycan peaks GP30, 36 and 38 showed the best discriminative power among individual N-glycans with C-statistics of:

- GP30 0.90 (0.85 - 0.96),
- GP36 0.87 (0.80 - 0.94),
- GP38 0.90 (0.83 - 0.96), (all $p \leq 8.4 \times 10^{-8}$, Figure 5.3.3 B, Table 5.3.3).

The best cut-off was chosen based on corresponding sensitivity and specificity from the ROC curve analysis (Table 5.3.3). The best performing GP30 had sensitivity of 89% and specificity of 79% when the cut-off <0.70% was applied; GP36 had sensitivity of 83% and specificity of 77% for the cut-off <1.61% and finally GP38 had sensitivity of 89% and specificity of 77% for the cut-off <0.25%.

The ROC curve analysis of hsCRP showed a satisfactory performance in discriminating between subjects with damaging/likely damaging *HNF1A* alleles and the participants without the *HNF1A* variants with AUC of 0.80 (0.69 - 0.90), $p=1.6 \times 10^{-5}$ (Figure 5.3.3 A). The CRP cut-off <0.8mg/L provided sensitivity of 78% and specificity of 69%.

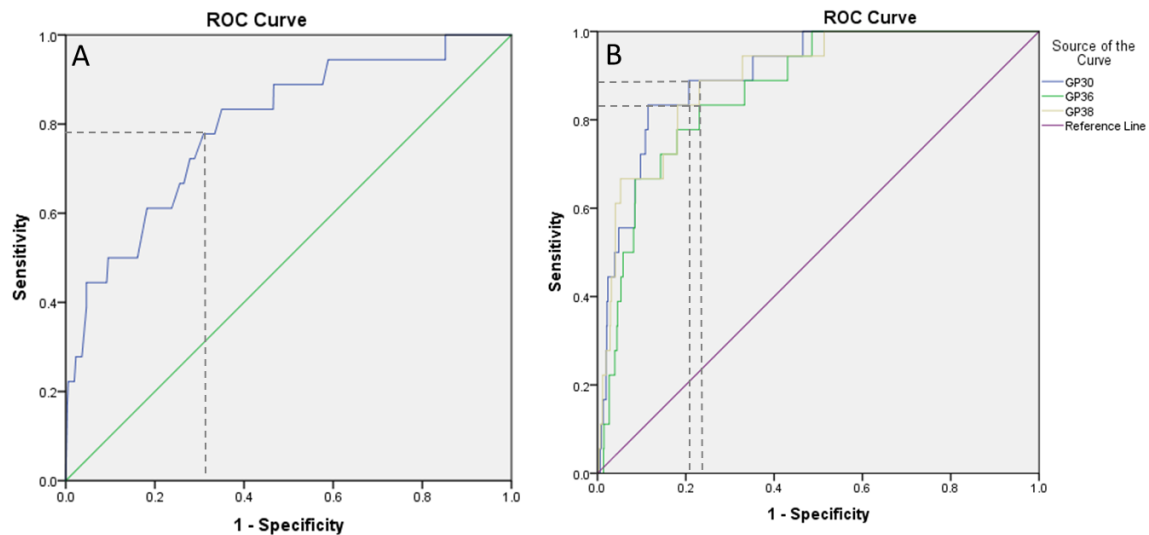


Figure 5.3.3 Receiver operating characteristic (ROC) curves depicting performance of hsCRP (A) and GP30, GP36 and GP38 (B) as a screening tool for *HNF1A*-MODY; C-statistics with AUC (area under the curve), cut-off values, sensitivity and specificity included in Table 5.3.3.

Biomarker	AUC	95% CI	p value	Cut-off	Sensitivity	Specificity
GP30	0.90	0.85-0.96	4.1×10^{-9}	<0.58%	83%	89%
				<0.70%	89%	79%
				<0.86%	94%	65%
GP36	0.87	0.80-0.94	8.4×10^{-8}	<1.50%	78%	82%
				<1.61%	83%	77%
				<1.83%	89%	67%
GP38	0.90	0.83-0.96	9.2×10^{-9}	<0.24%	83%	79%
				<0.25%	89%	77%
				<0.28%	94%	73%
hsCRP	0.80	0.69-0.90	1.6×10^{-5}	<0.70mg/L	72%	72%
				<0.80mg/L	78%	69%
				<1.00mg/L	83%	65%

Table 5.3.3 Area under the curve (AUC) characteristics for variables significantly lower in subjects with damaging/likely damaging *HNF1A* alleles versus those with no *HNF1A* variants; CI, confidence interval; three different cut-off values for each biomarker provided with corresponding sensitivity and specificity, in bold the chosen cut-off.

5.3.4 Clinical potential of GP30 and hsCRP

ROC curve analysis indicated that GP30 performed best in discriminating HNF1A-MODY from young-onset T2D, so its clinical potential was evaluated further on this study sample of individuals with non-autoimmune diabetes diagnosed before 45 years of age. HsCRP was also assessed due to its wide availability. A diagnostic threshold of 0.70% for GP30 provided good discrimination of HNF1A-MODY and early-onset T2D, with 89% sensitivity and 79% specificity. When GP30 was used as a screening tool for genetic testing on the current sample of individuals with non-autoimmune diabetes diagnosed before 45 years of age (n=977), 217 subjects (22.2%) with young-onset non-autoimmune diabetes would have been sequenced for *HNF1A*, resulting in the identification of 16 subjects with damaging/likely damaging *HNF1A* variants. Thus, two individuals with HNF1A-MODY would be missed (Table 5.3.4). The positive and negative predictive value (PPV and NPV) of the test was calculated and the PPV of GP30 was 7%. However, in the case of a screening test, high NPV is more important as we do not want to miss too many cases of HNF1A-MODY, and a definitive test (DNA sequencing) would be performed later on biomarker selected subjects. GP30 had the NPV of 99.7%.

According to an analysis of the ROC curve for hsCRP, a threshold of 0.80 mg/L provided sensitivity of 78% and specificity of 69%. The proposed CRP cut-off would lead to *HNF1A* sequencing of 279 subjects (30.2%) with early-onset non-autoimmune diabetes and the identification of 14 individuals with HNF1A-MODY. Hence, four patients with likely HNF1A-MODY would be missed. The NPV of CRP was similar to GP30 (99.4%) (Table 5.3.4).

The combination of both biomarkers (hsCRP cut-off <0.8mg/L and GP30 <0.70%) resulted in selecting 17 subjects with damaging/likely damaging *HNF1A* alleles. Thus, this approach would provide sensitivity of 94% but specificity of 57%. Using both biomarkers will select 431 subjects (44.1%) for genetic testing and have NPV of 99.8% similarly to GP30 or hsCRP.

	hsCRP (n=913)		GP30 (n=977)		hsCRP +GP30 (n=977)	
	All positive	All negative	All positive	All negative	All positive	All negative
Positive in test	14	275	16	201	17	414
Negative in test	4	620	2	758	1	545
Cases selected by test	289		217		431	
Sensitivity	0.78		0.89		0.94	
Specificity	0.69		0.79		0.57	
PPV	0.05		0.07		0.04	
NPV	0.99		1.00		1.00	
PLR	2.53		4.24		2.19	
NLR	0.32		0.14		0.10	

Table 5.3.4 Sensitivity, specificity, positive and negative predictive value (PPV and NPV), positive and negative likelihood ratio (PLR and NLR) of using hsCRP and GP30 as a screening tool for selecting individuals with likely *HNF1A*-MODY from the sample of individuals with non-autoimmune diabetes diagnosed before 45 years of age.

5.3.5 GP30 and CRP in predicting the functional effect of *HNF1A* variants

Both biomarkers (GP30 and hsCRP) were significantly lower in individuals with *HNF1A* alleles assigned damaging/likely damaging functional effect than in those with *HNF1A* alleles assigned likely benign/benign effect. Median GP30 was 0.43 (0.34-0.57) % in subjects with damaging/likely damaging *HNF1A* alleles versus 1.00 (0.64-1.29) % in those with likely benign/benign alleles ($p=1.2 \times 10^{-5}$). Median hsCRP was 0.21 (0.04-0.75) mg/L versus 1.40 (0.57-3.64) mg/L, $p=0.003$, when the same groups were compared (Figure 5.3.4).

The median hsCRP, GP30, GP36 and GP38 in subjects with likely benign/benign *HNF1A* alleles did not differ from the corresponding median biomarker level in individuals without *HNF1A* variants ($p=0.60-0.93$, Figure 5.3.4).

The hsCRP correlated negatively with the damaging functional effect of the *HNF1A* allele (Pearson's correlation coefficient $\rho=-0.45$, $p=0.006$). This reflected a continuum of phenotypical effect: median hsCRP 2.36mg/L for benign *HNF1A* alleles, 1.25mg/L for likely benign, 0.6mg/L for likely damaging and CRP 0.1mg/L for damaging *HNF1A* alleles (Figure 5.3.4A). A similar correlation applied to the median GP30 ($\rho=-0.57$, $p=3.9 \times 10^{-5}$), GP36 ($\rho=-0.61$, $p=3.1 \times 10^{-4}$) and GP38 ($\rho=-0.64$, $p=9.4 \times 10^{-5}$).

GP30 and hsCRP levels were analysed in subjects with rare non-synonymous *HNF1A* variants found in this study to test the hypothesis that biomarkers would be helpful in assigning a functional effect to these alleles (Table 5.3.5). Both biomarkers were in agreement with the allelic functional effect from the previous systematic assessment in the majority of subjects with rare *HNF1A* alleles (27/35), they were discordant (inconclusive) in 8/35 individuals (including 4 with damaging alleles: exon 1 del, p.G292fs, p.P379R and p.P379H) (Table 5.3.5, Figure 5.3.5). CRP and GP30 suggested opposite functional effects to the previously assigned ones in 4/35 subjects (with p.G151S, p.K222N and 2 individuals with p.A251T).

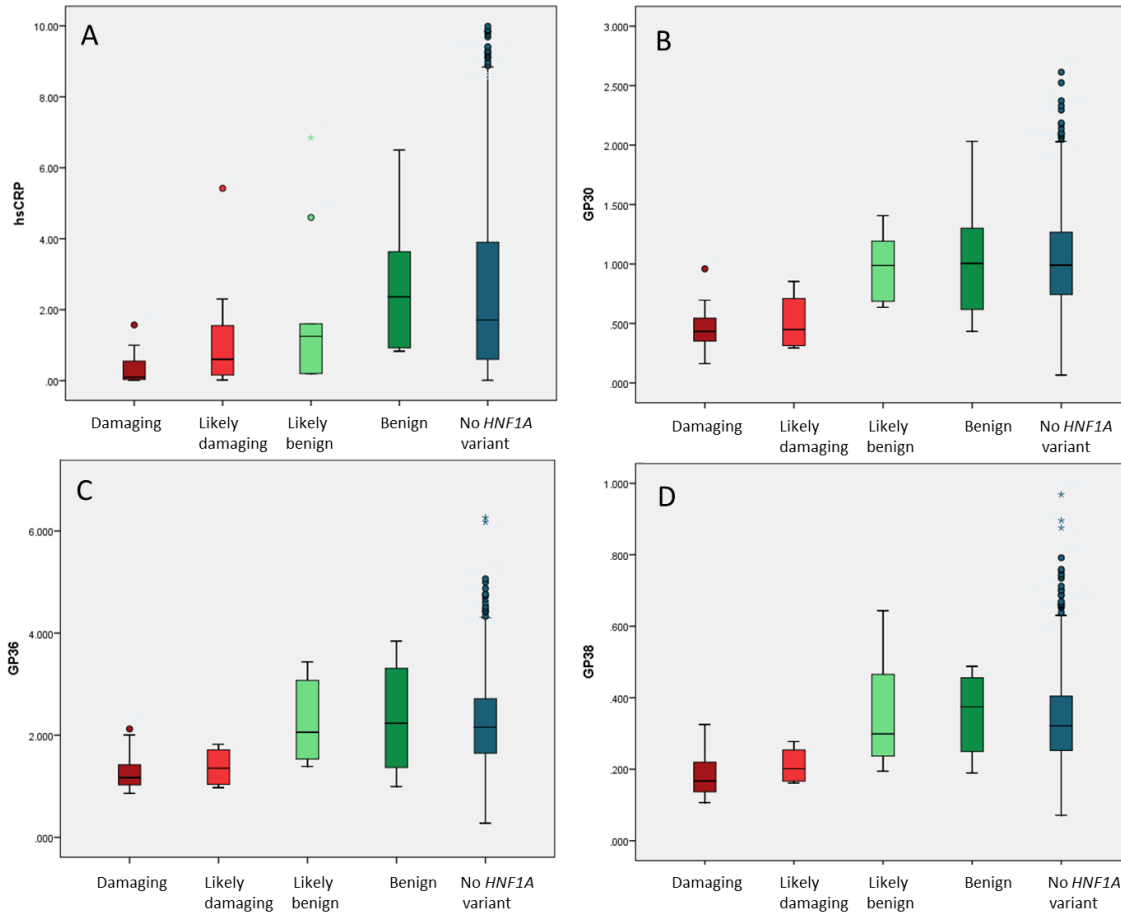


Figure 5.3.4 Median hsCRP (A), GP30 (B), GP36 (C) and GP38 (D) in subjects with rare HNF1A variant ($n=35$) divided in five groups depending on the predicted effect of HNF1A allele and in subjects without HNF1A variant ($n=878$ for hsCRP, $n=942$ for GP30); dark red, damaging alleles, medium red, likely damaging alleles; pale green, likely benign alleles; dark green, benign alleles; teal, no HNF1A rare variants. All biomarkers separate damaging/likely damaging alleles from all others; p value between subjects with damaging vs no HNF1A variants $<1 \times 10^{-7}$, p non-significant between groups with benign vs no HNF1A variants.

GP30 and hsCRP levels suggested that allele p.K222N has a damaging functional effect and p.G151S benign effect but with borderline levels of both markers. The subject with variant p.K222N presented with diabetes at 18 years of age, was sensitive to a small dose of sulphonylurea, had three generations' history of diabetes but was obese (BMI 32 kg/m^2). The individual with variant p.G151S presented with diabetes at 35 years of age,

had BMI of 26.3kg/m² ten years after diagnosis and had three generations' history of diabetes. Functional studies supported the damaging functional effect of allele p.G151S only. Borderline biomarkers, age at diagnosis and BMI in subject with p.G151S; and high BMI of the individual with p.K222N contribute to the discordant results and place these participants in the middle of the phenotypical spectrum of non-autoimmune diabetes. This shows again the complexity of assigning disease causality to genetic variants.

Variant p.A251T was present in three individuals in the present study. Two out of three subjects had both biomarkers at the level suggestive of a damaging functional effect, when the third subject had only CRP supporting a damaging effect of this allele. Variant p.A251T was assigned a likely benign functional effect with the systematic assessment discussed in chapter 3 which is consistent with clinical features of the proband but would be challenged by biomarker data.

Two individuals with PTVs (del Exon 1 and p.P292fs), subjects with variant p.P379R and p.P379H had discrepant biomarker results with one of the biomarkers (GP30) suggestive of a damaging functional effect of these alleles and all other evidence suggesting that those alleles are disease-causing (Table 5.3.5, Figure 5.3.5). In three of these subjects, GP30 was below 0.70% threshold and hsCRP was elevated which may suggest that CRP was affected by co-existing inflammatory process unrelated to HNF1A function.

Study ID	Protein variant	hsCRP	GP30	Previously predicted effect	Biomarker predicted effect
0617_003	P291fs	0.110	0.162	Likely damaging	Damaging
0957_003	P379fs	0.060	0.267	Damaging	Damaging
MCRO_055	P379R	0.600	0.294	Likely damaging	Damaging
2566_003	K222N	0.200	0.334	Likely benign	Damaging
2162_003	T260M	0.010	0.344	Damaging	Damaging
MCRO_294	R229*	0.400	0.351	Damaging	Damaging
2192_003	T260M	0.01	0.367	Damaging	Damaging
0555_003	Exon1del	5.420	0.376	Damaging	Inconclusive
MCRO_380	P379fs	0.100	0.422	Damaging	Damaging
2024_003	G339S	0.840	0.434	Benign	Inconclusive
0967_003	G292fs	0.100	0.445	Damaging	Damaging
2047_003	c.-4A>G	2.730	0.515	Benign	Inconclusive
MCRO_379	G292fs	0.400	0.519	Damaging	Damaging
2175_003	D135fs	0.010	0.522	Damaging	Damaging
0505_003	L377fs	0.210	0.545	Damaging	Damaging
MCRO_385	P379R	2.300	0.565	Likely damaging	Inconclusive
0785_003	P291fs	0.020	0.577	Damaging	Damaging
2550_003	L389V	12.070	0.617	Benign	Inconclusive
MCRO_245	A251T	0.300	0.635	Likely benign	Damaging
MCRO_345	A251T	0.200	0.643	Likely benign	Damaging
2196_003	P379H	1.570	0.695	Damaging	Inconclusive
MCRO_496	S3C	1.400	0.729	Likely benign	Benign
MCRO_042	G151S	0.800	0.854	Damaging	Benign
MCRO_343	A251T	0.200	0.877	Likely benign	Inconclusive
0587_003	E48K	2.000	0.953	Benign	Benign
0962_003	G292fs	0.700	0.959	Damaging	Inconclusive
2035_003	R583Q	3.29	1.004	Benign	Benign
0699_003	P291T	6.850	1.098	Likely benign	Benign
MCRO_072	T515M	4.600	1.106	Likely benign	Benign
MCRO_083	G288W	15.700	1.243	Benign	Benign
0522_003	P291T	1.600	1.277	Likely benign	Benign
MCRO_337	G288W	6.500	1.302	Benign	Benign
0561_003	P291T	14.760	1.407	Likely benign	Benign
2114_003	T196A	0.830	1.791	Benign	Benign
2510_003	H349Q	1.010	2.032	Benign	Benign

Table 5.3.5 Using hsCRP and GP30 level in assigning functional effect of HNF1A variant alleles. Variants of the study subject (including relatives) ordered from the smallest to the highest GP30; novel variants and variants with different prediction by both biomarkers when compared to the prediction of previous systematic assessment shaded in grey; all results supporting a damaging effect of the allele marked in bold.

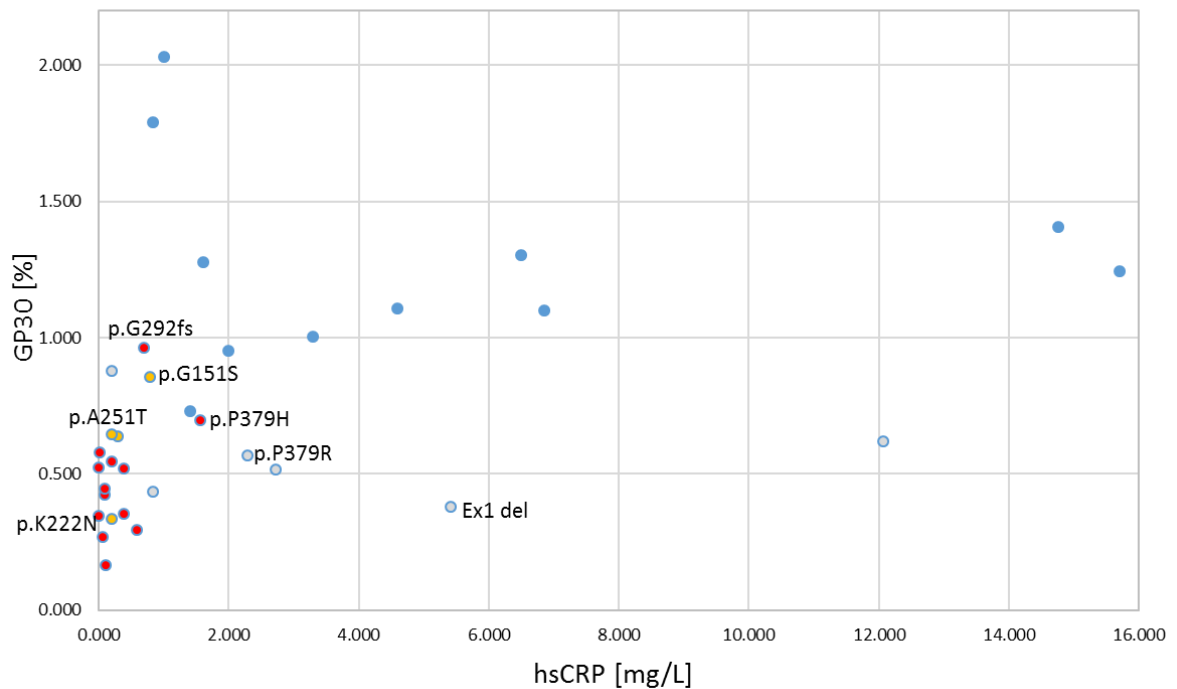


Figure 5.3.5 HsCRP [axis x] plotted against GP30 [axis y] for subjects with rare HNF1A alleles in the presented study. Damaging alleles marked in red, benign in blue, alleles with previously assigned functional effect opposite to the one suggested by biomarkers in orange and labelled (p.G151S, p.K222N and p.A251T), alleles with inconclusive biomarkers prediction in grey (damaging alleles with inconclusive biomarkers also labelled).

5.3.6 Clinical and biochemical predictors of HNF1A-MODY

I used a binary logistic regression to analyse factors associated with HNF1A-MODY to determine which variables select subjects with damaging/likely damaging HNF1A alleles (predict HNF1A-MODY) best. Variables were selected on the basis of their clinical significance, and included age at the time of diagnosis, gender, BMI, HbA1c, C-peptide, duration of diabetes, hsCRP, GP30, total cholesterol, triglycerides (TG), HDL and treatment modality (3 groups: insulin $\pm$ OHA, OHA and diet). Association of the selected variables

with presence of the disease was tested using univariate analysis. All variables with $p < 0.1$ on univariate analysis were included in multivariate model. As GP30 and C-peptide had very small values (most $< 2\%$ for GP30, $\text{hsCRP} < 2\text{mg/L}$ and $< 2\text{nmol/L}$ for C-peptide) and increase by 1 measure unit ($\%$ for GP30, mg/L for CRP and nmol/L for C-peptide) resulted in disproportional increase of odds ratio (OR), GP30 and C-peptide were transformed multiplying by 10 and hsCRP was Log_{10} transformed to normalise the data. Individuals with $\text{CRP} > 10\text{mg/L}$ were excluded as from previous analysis.

Univariate analysis showed that age at the time of diagnosis ($p < 0.0001$), GP30 ($p < 0.0001$), C-peptide ($p = 0.001$), hsCRP ($p = 0.002$), BMI ($p = 0.001$), TG ($p = 0.02$) and treatment modality ($p = 0.027$) were associated with diagnosis of HNF1A-MODY (Table 5.3.6).

Age at the time of diagnosis (OR 1.15, $p = 1.1 \times 10^{-4}$), C-peptide (OR 1.33, $p = 0.018$) and GP30 (OR 1.79, $p = 7 \times 10^{-6}$) were found to be independent predictors of the disease using multivariate logistic regression analysis. HsCRP remained in the model with OR of 1.55 but non-significant p value ($p = 0.08$, Table 5.3.6).

However, due to a small number of subjects in the group with damaging/likely damaging *HNF1A* alleles and disproportion of subject numbers between this group and group of individuals without *HNF1A* variants, the above model predicted correctly 98% of the cases without disease but only 2 subjects with HNF1A-MODY.

Variable	Univariable p	Multivariable analysis		
		log OR	OR	p value
Age Diagnosed	1.4×10^{-4}	0.136	1.146	1.1×10^{-4}
DM duration	0.718			
BMI	0.001			
Treatment	0.027			
Rx(Insulin)	0.01			
C-peptidex10	0.001	0.283	1.327	0.018
HbA1c	0.686			
T.cholesterol	0.841			
HDL	0.906			
TG	0.02			
hsCRP	0.002	0.439	1.551	0.08
GP30x10	2.6×10^{-5}	0.583	1.791	7.0×10^{-6}

Table 5.3.6 Univariate and multivariate analysis of factors associated with damaging *HNF1A* alleles (*HNF1A*-MODY) performed using data from 17 subjects with damaging/likely damaging *HNF1A* alleles versus 942 subjects with no *HNF1A* variants, only 4 most significant variables included in multivariate analysis (age at diagnosis, C-peptidex10, hsCRP and GP30x10); OR, odds ratio; TG, triglycerides.

5.4 Discussion

I have shown that antennary fucosylated plasma glycans (GP30, GP36 and GP38) and hsCRP were significantly lower in subjects harbouring damaging/likely damaging *HNF1A* alleles when compared to participants without *HNF1A* variants, and that biomarkers could assist with an initial classification of *HNF1A* variants. The 3-antennary fucosylated glycans performed well in differentiating subjects with likely damaging *HNF1A* alleles (likely *HNF1A*-MODY), from those without *HNF1A* variants (early onset T2D) with C-statistics of 0.87-0.90. High-sensitivity CRP performed satisfactorily in the same discrimination with C-statistic of 0.80.

This is consistent with known effects of HNF1A on plasma protein fucosylation and CRP, and previous findings (Thanabalasingham, Shah et al. 2011, Thanabalasingham, Huffman et al. 2013).

I performed a univariate analysis, in which CRP was a weaker and GP30 stronger independent predictor of HNF1A-MODY than age at diagnosis of diabetes. The weaker performance of hsCRP than age at diagnosis conforms to the findings of Bellanne-Chantelot *et al.* (Bellanne-Chantelot, Coste et al. 2016). In their study however, an excellent performance of clinical features alone (C-statistic 0.99) resulted from an over-fitted model with highly selected HNF1A-MODY patients (median BMI 23kg/m²) recruited from a tertiary centre. This diabetic population sample did not reflect the usual young adult diabetes clinic individuals.

Thus, the novelty of the results described in this chapter study was examining the performance of N-glycans and hsCRP in a relatively unselected population (only those with T1D excluded) of subjects with young adult-onset non-autoimmune diabetes, rather than well-defined groups of individuals with strong clinical features, as in the previous studies (Thanabalasingham, Shah et al. 2011, Thanabalasingham, Huffman et al. 2013, Bellanne-Chantelot, Coste et al. 2016). The subjects included in the current study came from consecutive recruitment without knowing their type of diabetes to exclude any bias towards MODY. This study approach better reflects the situation which clinicians encounter in young adult diabetes clinics or primary care when assigning a diagnosis. The distribution of age at diagnosis in my study subjects was comparable in both countries of

recruitment, while median BMI was slightly lower in subjects from Croatia but with a good spread of distribution (Chapter 3, Figure 3.2.2). Investigating less selected patient groups provides more accurate information regarding the prevalence of a rare disease, which was discussed in chapter 3, and excludes spectrum bias (Whiting, Rutjes et al. 2004).

Using a relatively unselected approach, 3-antennary fucosylated glycans (GP30, 36 and 38) provided similar C-statistics to DG9 glycan index in the previous study (Thanabalasingham, Huffman et al. 2013) discriminating HNF1A-MODY from T2D (C-statistics 0.87-0.90 vs previous 0.91). It might have been expected that the performance of biomarkers would be reduced in the unselected approach to the recruitment but these results could be explained by an improved method of N-glycan separation in this study (UPLC described in chapter 2). The unselected approach for the analysis of hsCRP performance in discriminating HNF1A-MODY from T2D had a satisfactory C-statistic but worse than in the previous study (0.80 vs previous 0.91) (Thanabalasingham, Shah et al. 2011) and consistent with a reduction in performance when performing unselective recruitment.

N-glycan analysis was performed with an improved methodology by using HILIC-UPLC instead of previously used HILIC-HPLC. The new method was validated and proved reliable and reproducible in quantitative glycan analysis (Huffman et al. 2014). HILIC-UPLC enabled a better separation of released N-glycans (38 glycan groups instead of 16). This allowed a direct employment of the initial individual glycan profile, without the need for the

desialylation of released glycans, which had to be performed previously. This reduced costs and time of the analysis.

Both GP30 and hsCRP performed well as screening tests and with a threshold of less than 0.70% and 0.80mg/L, respectively, picked up 89% and 78% of patients with likely disease-causing *HNF1A* alleles. Using both markers resulted in sensitivity of 94% but at the expense of lower specificity. The current diagnostic criteria for MODY miss many cases through low sensitivity (Shields, Hicks et al. 2010, Thanabalasingham, Pal et al. 2012), hence *HNF1A*-MODY remains misdiagnosed in the vast majority of cases (Shields, Hicks et al. 2010). If we use the classic clinical criteria for genetic testing (an age at diagnosis of diabetes younger than 25 years, C-peptide in the normal range and at least 2 generations family history of diabetes) we would pick up only 8 of 18 patients (44.4%) with the damaging/likely damaging *HNF1A* alleles identified in this study. Including biomarkers in the diagnostic pathway, such as CRP or GP30 which have high sensitivity, should result in an increase of the diagnostic rate of MODY.

Large scale sequencing studies in both healthy and diseased populations have shown that many variants, initially thought to be disease-causing, are found in population samples more often than expected for a rare monogenic condition (Flannick, Beer et al. 2013, Walsh, Thomson et al. 2017). GP30 and hsCRP could provide additional value in assigning disease causality of *HNF1A* alleles due to the fact that individuals with inactivating *HNF1A* variants had a significantly lower level of 3-antennary fucosylated glycans and hsCRP than subjects with benign *HNF1A* alleles.

HsCRP and GP30 were in agreement in assigning a functional effect to the majority of the *HNF1A* alleles. In the case of two novel variant alleles (p.G151S and p.K222N), biomarkers supported the disease causality of variant p.K222N and borderline benign effect of variant p.G151S. This prediction is opposite to the one from a systematic assessment of other features described in chapter 3 and shows that these biomarkers do not always agree with what the other features suggest about the allelic functional effect. However, this applies to only few alleles. In the case of p.G151S, borderline biomarkers, BMI and age at diagnosis place the individual with this allele in the middle of the phenotypical spectrum, which supports the fact that binary classification of MODY or T2D is no longer appropriate.

The median hsCRP, GP30, GP36 and GP38 negatively correlated with the increasing damaging functional effect of *HNF1A* alleles (from benign, through likely benign, likely damaging to damaging) and also reflected the phenotypical spectrum of the genetic variation rather than a binary outcome (*HNF1A*-MODY or T2D).

CRP is an inexpensive and widely available test, and despite its slightly poorer performance in selecting patients with *HNF1A*-MODY, it has immediate translational potential for selecting individuals for genetic testing and should be included in diagnostic algorithm for a young adult onset diabetes. However, the disadvantage of using hsCRP is that it increases during the inflammatory response and hence, if significantly elevated, would require the test to be repeated at a later stage.

Regarding the N-glycan profile analysis, the main obstacles are the high price and limited availability of the test thus GP30 cannot currently be included in a clinical algorithm. Thus, a design of a simple diagnostic method such as ELISA that measures 3-antennary fucose only could significantly reduce the price of the test.

5.5 Conclusions

In conclusion, GP30 and hsCRP performed well in discriminating subjects with damaging/likely damaging *HNF1A* alleles from individuals with early onset T2D and no rare *HNF1A* variants. A high NPV and C-statistics of both biomarkers make them good candidates for screening tests of HNF1A-MODY. A diagnostic protocol combining clinical features with biochemical markers should improve the diagnostic rate of HNF1A-MODY and clinical management of this subtype of diabetes. Furthermore, GP30 and hsCRP could contribute to assigning a functional effect of *HNF1A* alleles and should be used for that purpose if available.

Chapter 6

Urinary steroid profile in HNF1A-MODY patients

The contents of this chapter have been presented at the British Endocrine Society conference:

- Juszczak A, Gilligan LC, Hughes BA, Hassan-Smith ZK, Arlt W, McCarthy MI, Tomlinson JW, Owen KR. *Distinct patterns of glucocorticoid metabolism in patients with HNF1A-MODY have the potential to confer a beneficial metabolic phenotype. Endocrine Abstracts (2017) Vol 50, p54-55.*

6.1 Introduction

HNF1A is a transcription factor for around 220 genes and is therefore involved in multiple metabolic pathways. As I discussed in chapter 1, section 1.6, potential biomarkers of HNF1A-MODY, such as apolipoprotein M, CRP or N-glycan profile, have been investigated due to the fact that HNF1A was known to regulate expression of their genes or enzymes involved in their metabolism. Hence, I reviewed published literature for any other putative biomarkers of HNF1A-MODY and their relevance to the pathogenesis of this type of diabetes.

It has been found that several steroidogenesis genes were downregulated in the liver cells of *Hnf1a* knock-out mice (Servitja, Pignatelli et al. 2009) and thus this chapter focuses on urinary steroid metabolites and enzymes involved in steroid metabolism.

6.1.1 Steroid metabolism genes downregulated in the hepatocytes of *Hnf1a* knock-out mice

Genes encoding for several enzymes in steroid metabolism are listed in chapter 1, section 1.7.4. This includes the genes coding for 11 β -hydroxysteroid dehydrogenase type 1 (*Hsd11b1*), converting inactive cortisone to active cortisol, and 3 β -hydroxysteroid dehydrogenase type 2 (*Hsd3b2*) which catalyses one of the initial steps in the synthesis of all adrenal steroids. The RNA expression of *Hsd11b1* in the hepatocytes of the *Hnf1a* $-/-$ mice was downregulated 15-fold and *Hsd3b2* 35-fold when compared to the wild type (WT) *Hnf1a*. Additionally, aldo-keto-reductase family 1 member 6 (*Akr1c6*) involved in the conversion of steroids was downregulated 17-fold.

6.1.2 11 β -hydroxysteroid dehydrogenases and their involvement in glucose metabolism

In humans, 11 β -hydroxysteroid dehydrogenase type 1 (HSD11B1) interconverts inactive cortisone to active cortisol with predominant reductase activity (cortisone to cortisol) (Tomlinson, Walker et al. 2004). HSD11B1 is expressed at RNA and protein level mainly in the liver, less in the skeletal muscles, adipose tissue and very low expression in adrenal glands. 11 β -hydroxysteroid dehydrogenase type 2 (HSD11B2) catalyses the reaction deactivating cortisol to cortisone exclusively (Figure 1.6.2) and its mRNA and protein is expressed in the mineralocorticoid target tissues, mainly the kidneys, GI tract and placenta (Ricketts et al. 1998, Uhlen, Fagerberg et al. 2015).

Several studies aimed to explain pathogenesis of type 2 diabetes by investigating 11BHSD1 or other enzymes involved in glucocorticoid metabolism. However, there is no literature describing the activity of the enzymes involved in steroid metabolism in subjects with HNF1A-MODY.

An increased global activity of HSD11B1 was thought to have a role in the pathogenesis of obesity and metabolic syndrome (Bujalska et al. 1997), however it was proven that 11 β -HSDs work on a tissue-specific level (Rask et al. 2002). Subsequently, Valsamakis *et al.* showed that there was no difference in HSD11B1 activity based on urinary steroid levels in obese men with T2D (n=33) when compared to BMI-matched non-diabetic subjects (n=38) (Valsamakis et al. 2004).

Tomlinson *et al.* studied 101 obese subjects (including 66 women and 45 men, mean BMI 34.4kg/m²) and reported that 22.7% of obese women and 34.2% of obese men in this study had impaired glucose tolerance and 7 of them were diagnosed with diabetes (Tomlinson *et al.* 2008). An impaired glucose tolerance and insulin resistance were associated with an increased expression of *HSD11B1* mRNA in the adipose tissue but only in women. Nonetheless, the prevalence of T2D is not different between genders which would suggest that activity of 11 β -HSD type 1 does not contribute to pathogenesis of T2D.

The selective inhibition of HSD11B1 in hyperglycaemic hyperinsulinaemic mice (ob/ob and KKAy) resulted in a significantly improved oral glucose tolerance (glucose reduced by 25-35% when compared to untreated animals) and insulin sensitivity (insulin reduced by 50% when compared to untreated animals), making it a good candidate for diabetes treatment (Alberts, Engblom *et al.* 2002, Alberts, Nilsson *et al.* 2003). However, phase II studies of selective inhibitors of HSD11B1 in obese human subjects with T2D showed none or only a modest effect on hyperglycaemia and body weight (FPG reduced by 0.3-1.3mmol/L, HbA1c decreased by 0.56%, body weight reduced by 1-1.8kg) (Rosenstock *et al.* 2010, Feig *et al.* 2011). Hence, data from studies on HSD11B1 inhibitors in humans again did not support a role of 11 β -HSD type 1 in T2D.

Cortisol can also be deactivated to 5 α - or 5 β -metabolites by 5 α - and 5 β -steroid reductases (SRD5A, SRD5B) (Russell and Wilson 1994) and may compensate for the effect on 11 β -HSD type 1 on cortisol. There are 3 isoforms of SRD5A protein, including: SRD5A1, which is most abundant in humans and expressed mainly in liver and adipose tissue; and

SRD5A2, which is mainly expressed in the liver and genital tract. SRD5B (AKR1D1) is expressed in the liver with low expression in the testes and Fallopian tubes (Uhlen, Fagerberg et al. 2015). The *Srd5a1* knock-out mice developed insulin resistance and liver steatosis on an obesogenic diet (Dowman et al. 2013). An elevated global activity of SRD5A with increased 5 α -metabolites of cortisol have also been reported in obese human subjects with biopsy-proven fatty liver (n=8) when compared to BMI-matched subjects (n=32) (Ahmed et al. 2012).

Selective inhibition of SRD5A2 with Finasteride but not Dutasteride (inhibitor of SRD5A1 and 2) showed a potential benefit on the human metabolic phenotype. Hazlehurst *et al.* performed a randomised study to examine an effect of 21 days' treatment with one of the above SRD5A inhibitors on liver fat deposition in 12 men (Hazlehurst et al. 2016). Men treated with Dutasteride but not Finasteride had increased fat deposition in the liver proven by hepatic magnetic resonance spectroscopy, as well as increased lipogenesis in the adipose tissue measured by adipose microdialysis. This implies that the inhibition of SRD5A1 may contribute to the metabolic syndrome in humans and therefore T2D.

The above studies imply that a changed activity of steroid reductases could contribute to the phenotype of T2D and may explain the phenotypical differences between T2D and HNF1A-MODY.

6.1.3 3 α - and 3 β -hydroxysteroid dehydrogenases

3 β -hydroxysteroid dehydrogenase type 2 (HSD3B2) catalyses one of the initial steps in the synthesis of all adrenal and gonadal steroids where it is exclusively expressed; HSD3B1 is expressed in the placenta and skin. HSD3B2 deficiency is a rare cause of classic congenital adrenal hyperplasia, which presents in childhood with salt wasting due to aldosterone and cortisol deficiency.

Four 3 α -hydroxysteroid dehydrogenases (aldo-keto-reductases, AKR1C1-4) are known to be present in humans (Qin et al. 1993), of which AKR1C4 is mainly expressed in the liver and catalyses the second step of deactivation of cortisol and cortisone along with 5 α - and 5 β -steroid reductases. AKR1C1-3 is expressed in the liver, adrenals, testes and prostate and deactivates androstenedione and dihydrotestosterone.

Based on the data from the *Hnf1a* knock-out mice, I hypothesised that the activity of HSD11B1 and HSD3B2 would also be reduced in the haploinsufficient state of *HNF1A* in patients with HNF1A-MODY when compared to non-diabetic individuals or patients with T2D. The activity of HSD11B2, SRD5A and SRD5B was also of interest as these enzymes are involved in deactivation of cortisol and so are essential to understanding changes of HSD11B1 or HSD3B2. Altered steroid metabolism may explain why patients with HNF1A-MODY have a lean phenotype and normal insulin sensitivity when compared to individuals with type 2 diabetes. All 3 α -hydroxysteroid dehydrogenases are highly expressed in the human liver and adrenal glands and are not rate-limiting enzymes, thus they were not

studied. The urinary steroid profile of individuals with HNF1A-MODY has not been published in previous literature.

6.1.2 Aims of the study

In this study I aimed to:

1. Compare the urinary steroid profile in subjects with HNF1A-MODY, T2D and healthy controls;
2. Establish if the urinary steroid profile in HNF1A-MODY subjects can explain their lean phenotype and normal insulin sensitivity.

6.2 Study participants and methods

6.2.1 Study subjects

35 individuals with HNF1A-MODY, 35 with type 2 diabetes and 35 healthy controls were recruited and clinical information on age at diagnosis of diabetes, treatment history, current medications and family history of diabetes were recorded for subjects with diabetes. A diagnosis of HNF1A-MODY was based on the presence of *HNF1A* allele reported previously to cause HNF1A-MODY and co-segregating with diabetes in the family. The initial selection of individuals with young-adult onset diabetes for the *HNF1A* sequencing was based on the clinical features (negative GAD antibodies, C-peptide in the normal range despite more than three years from diagnosis of diabetes and sensitivity to sulphonylurea). The subjects with T2D were selected based on the clinical features (BMI>25kg/m², C-peptide in the mid- to upper normal range and features of metabolic

syndrome such as elevated triglycerides). Anthropometrical measurements were collected. Basic biochemical tests such as HbA1c, fasting plasma glucose, C-peptide, renal function and lipid profile were measured. Subjects with both types of diabetes were selected from the YDX (Young Diabetes in Oxford) study described in chapter 2, section 2.1.1. The healthy-control participants were selected from the Healthy Human Aging study which recruited from the local Birmingham population (Hassan-Smith et al. 2015) and was described in chapter 2, section 2.1.3. These subjects had their fasting plasma glucose measured to exclude undiagnosed diabetes and did not have a history of IHD, CVD, respiratory disease or epilepsy. Spot morning urine samples were collected from all diabetic individuals, while healthy controls provided 24 hours urine samples. Female subjects were not on any contraceptive or hormonal replacement therapy and individuals on any form of steroid treatment were excluded. One subject from the HNF1A-MODY group with serum creatinine above $160\mu\text{mol/L}$ was excluded.

Both studies had local ethics committee approval.

6.2.2 Study sample

The study sample was determined by the number of available subjects with HNF1A-MODY ($n=35$, including 18 women, 17 men) and a T2D and healthy control group were matched in sample size. Due to the gender dimorphism of many adrenal steroids in humans, the effect size was calculated separately for men and for women. The effect size calculations were based on the SD of the steroid ratios in the examined study group in each gender and calculated using the T-statistic and an online effect size calculator

(<http://www.sample-size.net/means-effect-size/>). To achieve a power of 80% with α of 0.05 with the above groups, the effect size was calculated as 0.58 for men and 1.19 for women for 5BTHF/5ATHF ratio; 0.37 and 0.75 for Etiocholanolone/androsterone ratio; 0.20 and 0.24 for cortisol/cortisone ratio respectively.

6.2.3 Methods

6.2.3.1 GCMS and derived urinary steroid ratios

The urinary steroid profile was measured in spot urine samples by GCMS as described in chapter 2, section 2.15, and the values being expressed in $\mu\text{g/L}$ of urine. The measurement of the urinary steroid profile was performed in the Laboratory of the Institute of Metabolism and Systems Research, University of Birmingham by Beverly Hughes and Lorna Gilligan.

Thirty-three different urinary steroids and their metabolites were separated by gas chromatography/mass spectrometry (GC/MS) on the Agilent 5975 instrument using the previously described method (Shackleton 1986, Krone, Hughes et al. 2010). In summary, steroid sulphate and glucuronide conjugates were first extracted from urine and chemically released from conjugates by enzymatic hydrolysis to make them stable at a high temperature. Subsequently, freed steroids were re-extracted and derivatised to form methyloxime-trimethyl-silyl ethers. Samples were injected sequentially and separated on increasing the temperature gradient (100-300°C). The individual steroids were ionised in the fused-silica column and specific ions formed from different metabolites. Internal

standards were used for quantification. Previously stored urine samples of healthy individuals were used for quality control.

Seventeen different steroid ratios were derived from different metabolites based on steroidogenesis pathway and steroid metabolism and those are used in the clinical practice for the interpretation of steroid metabolism (Figure 1.7.2). Many adrenal steroids express gender dimorphism, thus a comparison was made separately for men and women. Due to the fact that some crucial enzymes in steroidogenesis convert multiple substrates at different levels of the pathway, the ratios derived from steroid metabolites are regarded more relevant than single steroid measures (Wooding and Auchus 2013).

Based on the data from *Hnf1a* $-/-$ mice and other enzymes involved in cortisol metabolism, the ratio defining the activity of HSD11B1, HSD11B2, HSD3B2, SRD5A and SRD5B were selected as of interest. The function of HSD11B1 was estimated by the ratio of urinary metabolites of cortisol (F) to the metabolite of cortisone (E): $(5BTHF+5ATHF)/THE$, where 5BTHF is 5β -tetrahydrocortisol; THE, tetrahydrocortisone and 5ATHF, 5α -THF (Palermo et al. 1996) (Figure 6.2.1). The ratios of F/E and $F+E/(5BTHF+5ATHF+THE)$ were used to estimate the function of HSD11B2. The activity of the total HSD3B was expressed by dividing the substrate by the metabolites of the final product (cortisol), so $DHEA/(5BTHF+5ATHF+THE)$, where DHEA is dehydroepiandrosterone; (Figure 1.7.2).

The ratios of 5BTHF/5ATHF, and Etio/Andro defined the activity of the total SRD5A and SRD5B (where Etio is etiocholanolone, the final metabolite of 5β -THF; 5β -THE and 5β -

dihydrotestosterone; Andro, androsterone is the final metabolite of 5 α -THF and 5 α -dihydrotestosterone).

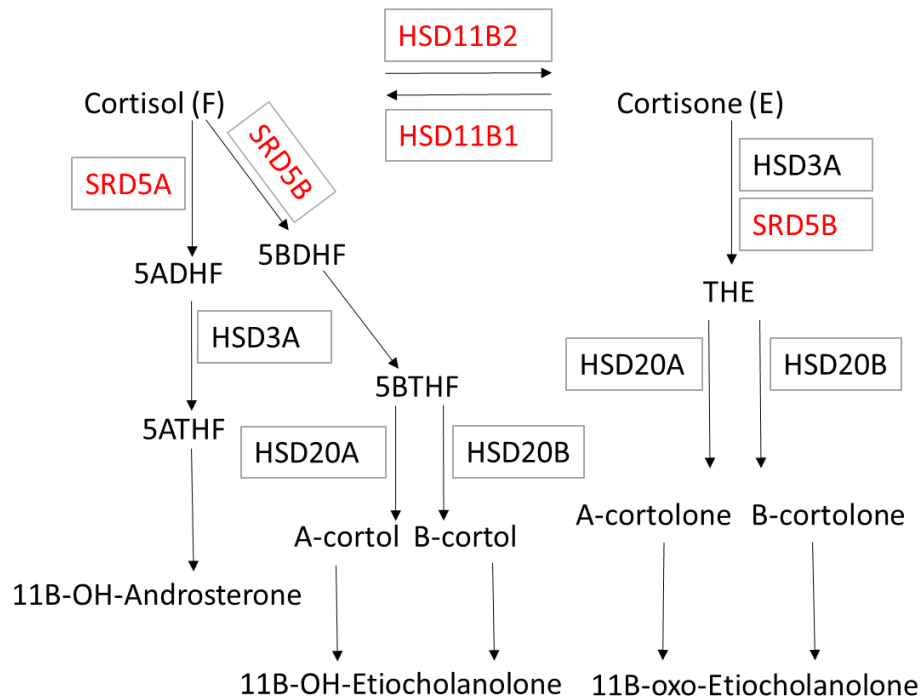


Figure 6.2.1 Metabolism of cortisol with enzymes involved presented in rectangles and the main enzymes discussed in this chapter in red. HSD11B1, 11 β -hydroxysteroid dehydrogenase type 1; HSD11B2, 11 β -hydroxysteroid dehydrogenase type 2; SRD5A, 5 α -steroid reductase; SRD5B, 5 β -steroid reductase; THF, tetrahydrocortisol; DHT, dihydrocortisol; THE, tetrahydrocortisone; Etio, etiocholanolone; Andro, androsterone; HSD3A, 3 α -hydroxysteroid dehydrogenase; HSD20A and 20B, 20 α - and 20 β -hydroxysteroid dehydrogenase.

6.2.3.2 Calculation of creatinine-corrected urinary steroid values

To calculate the sample size for investigating the correlation between creatinine-corrected and non-corrected steroid metabolites, I used the equation:

$$N = \left[\frac{Z\alpha + Z\beta}{C} \right]^2 + 3$$

$$C=0.5*\ln[(1+r)/(1-r)]$$

Where $Z_{\alpha}=1.96$ and is standard normal deviate for $\alpha=0.05$, $Z_{\beta}=0.84$ and is standard normal deviate for $\beta=0.2$ (type II error rate) and r is an expected correlation coefficient of 0.5. The required minimal sample size using this equation is 29.

Therefore, for the first 50 subjects in this study, including individuals with HNF1A-MODY and T2D, the urine creatinine levels were measured and the values of the individual urinary steroids were corrected for creatinine (as $\mu\text{g/g}$ of creatinine) using the following formula:

$$x = \frac{\text{uncorrected steroid } (\mu\text{g/L}) \times 100}{\text{creatinine } (\text{mg/dL})}$$

As the uncorrected steroid values correlated very well with the creatinine-corrected ones (described in results), the measuring of urine creatinine concentration was abandoned for the rest of the study participants.

6.2.3.3 Statistical analysis

IBM SPSS v23 was used for statistical analysis. Subject characteristics were expressed as a mean ($\pm$ SD) for continuous variables with a normal distribution, as a median (IQR) for those with a non-parametric distribution, and as percentages for categorical measures. Differences in the frequency of categorical variables were tested using the Chi-squared test or Fisher's exact test. For descriptive statistics, a p value of <0.05 was considered significant.

The majority of urinary steroid ratios had a non-parametric distribution, and thus were expressed as a median (IQR) and the p value was calculated using the Mann-Whitney U test when comparing HNF1A-MODY patients versus two other groups separately. Significant differences were considered for $p < 0.032$, which was calculated using the Benjamini-Hochberg procedure for a total of 50 metabolites/ratios analysed and an accepted false discovery rate of 10%. All $p < 0.05$ were ranked from the smallest to the highest and adjusted p value calculated using the formula:

$$p_{BH} = i/(m) \times Q$$

Where i denotes rank, m the number of different tests (hypotheses) investigated and Q the false discovery rate. This reduced the significant p value to <0.032 .

A significant correlation was defined by p value <0.05 using Pearson's correlation coefficient.

6.3 Results

6.3.1 Subject characteristics

All subjects were age- and BMI-matched in each gender group (Table 6.2.1). However, women with T2D had a higher mean waist-to-hip ratio when compared to those with HNF1A-MODY ($0.93 (\pm 2.27)$ vs $0.82 (\pm 0.06)$, $p = 4.0 \times 10^{-5}$) despite matching BMI in both groups. There was no difference in waist-to-hip ratio in men with either type of diabetes (Table 6.2.1).

The HbA1c was similar in both groups with diabetes, while the median duration of diabetes was longer in HNF1A-MODY subjects in both genders [11.3 (7.7-18.6) years in HNF1A-MODY versus 3.0 (1.3-13.0) years in T2D group in men, $p=0.008$] and [21.4 (8.8-31.0) years vs 6.4 (3.0-14.7) years in women respectively, $p=0.047$].

Subjects with HNF1A-MODY had significantly lower insulin resistance, measured using an HOMA2-IR calculator (Wallace et al. 2004), when compared to BMI-matched subjects with T2D in both genders [median 0.95 (0.56-1.28) vs 2.19 (1.69-2.80), $p=0.004$ in women and 1.37 (0.64-1.67) vs 1.63 (1.42-2.19), $p=0.04$ in men]. The number of patients treated with insulin or statin did not differ between HNF1A-MODY and T2D groups.

There was no significant difference in systolic and diastolic blood pressure (BP) in women in all three groups. Men with HNF1A-MODY had a higher systolic BP when compared to healthy controls [median 139 (130.5-148.5) mmHg vs 125 (117-128) mmHg, $p=0.02$] but there was no difference in systolic BP between both diabetes groups (Table 6.2.1). However, individuals with T2D were more often treated with anti-hypertensive medications (7/18 men, 12/18 women with T2D, 2 men and 2 women in HNF1A-MODY group).

	Study group			p value	
	HNF1A-MODY	T2D	Control	HNF1A vs T2D	HNF1A vs Control
Men					
Number of subjects	n=16	n=18	n=17		
Mean (SD) age [years]	39.43 (11.16)	38.49 (5.99)	38.41 (9.97)	0.83	0.85
Mean (SD) BMI [kg/m ²]	26.92 (4.61)	26.69 (3.17)	25.83 (2.56)	0.75	0.49
Mean (SD) Waist/hip ratio	0.91 (0.06)	0.91 (0.08)	n/a	0.88	n/a
Median (IQR) systolic BP [mmHg]	139 (130.5-148.5)	136 (121-139)	125 (117-128)	0.39	0.02
Median (IQR) diastolic BP [mmHg]	84 (72.5-89.5)	81 (76-85)	80 (72-85)	0.88	0.50
Mean (SD) creatinine [μ mol/L]	84.78 (13.82)	81.89 (23.79)	n/a	0.20	n/a
Mean (SD) K+ [mmol/L]	4.02 (0.47)	4.28 (0.57)	n/a	0.22	n/a
Median (IQR) onset of diabetes [years]	24.50 (20.50-33.50)	35.00 (28.00-38.00)	n/a	0.07	n/a
Median (IQR) DM duration [years]	11.26 (7.65-18.60)	2.96 (1.26-13.00)	n/a	0.01	n/a
Median (IQR) HbA1c [%]	7.20 (6.35-7.80)	6.70 (6.50-7.30)	n/a	0.81	n/a
Median (IQR) FPG [mmol/L]	6.70 (6.20-8.60)	7.20 (6.30-9.50)	5.10 (4.80-5.10)	0.56	1.4×10^{-7}
Median (IQR) HOMA2-IR	1.37 (0.64-1.67)	1.63 (1.42-2.19)	1.17 (0.71-1.91)	0.04	0.81
Mean (SD) total cholesterol [mmol/L]	4.80 (0.97)	4.60 (0.99)	4.76 (1.13)	0.47	0.74
Mean (SD) HDL [mmol/L]	1.29 (0.45)	1.02 (0.20)	1.27 (0.21)	0.10	0.41
Median (IQR) serum cortisol [nmol/L]	318.5 (260-383)	250 (211.5-354)	285 (236.5-373)	0.20	0.67
Insulin treated	5 (31.3%)	4 (22.2%)	n/a	0.24	n/a
Statin treated	4 (25%)	7 (38.9%)	n/a	0.31	n/a
HTN treated	2 (12.5%)	7 (38.9%)	n/a	0.07	n/a
Women					
Number of subjects	n=18	n=18	n=18		
Mean (SD) age [years]	44.96 (16.32)	48.42 (10.34)	42.17 (16.07)	0.67	0.61
Mean (SD) BMI [kg/m ²]	26.74 (2.58)	26.55 (2.27)	26.16 (3.38)	0.55	0.39
Mean (SD) Waist/hip ratio	0.82 (0.06)	0.93 (0.06)	n/a	4.0×10^{-5}	n/a
Median (IQR) systolic BP [mmHg]	128 (115.0-144.5)	132 (122-141)	123 (113-133)	0.57	0.42
Median (IQR) diastolic BP [mmHg]	76 (68-80.5)	78.5 (74-87)	74.5 (63-83)	0.22	0.69
Mean (SD) creatinine [μ mol/L]	74.26 (19.17)	82.83 (35.32)	n/a	0.91	n/a
Mean (SD) K+ [mmol/L]	3.89 (0.32)	4.37 (0.59)	n/a	0.01	n/a
Median (IQR) onset of diabetes [years]	19.50 (16.00-28.00)	40.00 (32.00-44.00)	n/a	9.0×10^{-6}	n/a
Median (IQR) DM duration [years]	21.38 (8.84-31.00)	6.40 (2.99-14.66)	n/a	0.05	n/a
Median (IQR) HbA1c [%]	7.00 (5.90-7.70)	6.90 (6.40-8.10)	n/a	0.54	n/a
Median (IQR) FPG [mmol/L]	6.05 (4.85-7.60)	7.90 (6.80-11.55)	4.55 (4.30-5.50)	0.03	0.001
Median (IQR) HOMA2-IR	0.95 (0.56-1.28)	2.19 (1.69-2.80)	0.96 (0.77-1.30)	0.004	0.80
Mean (SD) total cholesterol [mmol/L]	4.30 (0.61)	4.49 (1.15)	4.96 (0.79)	0.77	0.02
Mean (SD) HDL [mmol/L]	1.59 (0.46)	1.13 (0.30)	1.72 (0.59)	0.004	0.55
Median (IQR) serum cortisol [nmol/L]	282 (234.5-344.5)	260 (210.5-399)	286 (226-362)	0.87	0.64
Insulin treated	8 (44.4%)	7 (38.9%)	n/a	0.50	n/a
Statin treated	5 (27.8%)	10 (55.6%)	n/a	0.09	n/a
HTN treated	2 (11.1%)	12 (66.7%)	n/a	0.0007	n/a

Table 6.2.1 Patient characteristics; DM, diabetes mellitus; T2D, type 2 diabetes; FPG, fasting plasma glucose; BMI, body mass index. Mann-Whitney U test was used for comparing non-parametric variables and t-test for normally distributed ones in patients with HNF1A-MODY vs those with T2D or non-diabetic controls; p value <0.05 was considered statistically significant.

6.3.2 Urinary steroid metabolites in HNF1A-MODY

Table 6.3.1 and 6.3.2 summarise all altered urinary steroids and their metabolite ratios, and the ratios describing the activity of the enzymes in my study hypothesis in men and women. Multiple steroids and their metabolites were elevated in men with HNF1A-MODY when compared to healthy controls and included glucocorticoids such as cortisol (F) and its metabolites [6B-OH-cortisol, tetrahydrocortisol (THF), cortisone (E), tetrahydrocortisone (THE)] and mineralocorticoids such as tetrahydro-corticosterone (THB) and tetrahydro-11-deoxycortisol (THS), ($p=0.0002-0.03$, Table 6.3.1).

Free cortisol and 6B-OH-cortisol were also elevated in females with HNF1A-MODY in comparison to the age and BMI matched healthy controls ($p=0.016$ and 0.024 respectively). Cortisol and cortisone, but not their 5α - or 5β -metabolites, were also significantly elevated in both genders with T2D when compared to healthy controls ($p\leq 0.001$).

However, the median serum cortisol was not significantly different between the three groups in both genders (Table 6.2.1) suggesting again that differences in steroid metabolites are at the specific target tissue and not circulating level.

Urinary Steroids and their ratios	Median urinary steroid/ ratio (IQR)			p value*	
	HNF1A-MODY	T2D	Control	HNF1A vs T2D	HNF1A vs CTR
Men					
Individual steroids					
Etio (Etiocanolone)	1582.0 (1044.5-3306.0)	1155.0 (703.0-3490.0)	803.0 (641.0-1076.0)	0.40	0.02
Andro (Androsterone)	2041.0 (1099.5-3118.5)	2249.0 (1175.0-3247.0)	888.0 (755.0-1976.0)	0.62	0.08
5PD (pregnenediol)	461.5 (216.5-600.0)	577.0 (338.0-811.0)	122.0 (65.0-197.0)	0.14	1.7 × 10⁻⁴
Pregnadienol	200.5 (84.0-307.0)	241.0 (136.0-358.0)	57.0 (35.0-102.0)	0.38	0.03
THA (tetrahydro-11-dehydrocorticosterone)	90.5 (46.5-162.5)	71.0 (22.0-114.0)	38.0 (26.0-88.0)	0.30	0.03
THB (tetrahydro-corticosterone)	116.0 (53.5-230.5)	82.5 (22.0-177.0)	46.0 (24.0-88.0)	0.28	0.006
TH-DOC (tetrahydro-11-desoxycorticosterone)	14.0 (8.3-21.5)	8.5 (4.0-17.0)	6.0 (5.0-9.0)	0.06	0.001
PD (pregnanediol)	191.5 (118.5-310.5)	139.0 (70.0-229.0)	72.0 (56.0-106.0)	0.24	0.004
17HP (17-hydroxypregnanolone)	152.5 (93.5-248.0)	99.5 (49.0-138.0)	81.0 (35.0-91.0)	0.06	0.001
THS (tetrahydro-11-desoxycortisol)	58.5 (33.0-70.0)	35.0 (12.0-68.0)	28.0 (16.0-40.0)	0.16	0.002
F (cortisol)	131.5 (49.0-196.0)	100.5 (48.0-152.0)	26.0 (19.0-41.0)	0.67	1.5 × 10⁻⁴
6B-OHF (6beta-hydroxycortisol)	161.5 (83.5-277.0)	176.0 (40.0-241.0)	53.0 (40.0-65.0)	0.46	0.001
5BTHF (tetrahydrocortisol)	1881.0 (805.5-2209.5)	1218.5 (345.0-2177.0)	749.0 (478.0-1064.0)	0.14	0.004
5ATHF (5alpha-tetrahydrocortisol)	1054.5 (716.5-1937.0)	1036.0 (633.0-1528.0)	894.0 (530.0-1270.0)	0.60	0.35
E (cortisone)	144.0 (61.0-215.5)	135.5 (82.0-206.0)	37.0 (32.0-57.0)	0.85	2.5 × 10⁻⁴
THE (tetrahydrocortisone)	3196.0 (1297.5-3841.0)	2661.0 (978.0-4298.0)	1572.0 (928.0-2531.0)	0.46	0.02
Steroid ratios					
F/E	0.82 (0.73-1.05)	0.77 (0.71-0.89)	0.65 (0.59-0.78)	0.33	0.004
(F+E)/(5BTHF+5ATHF+THE)	0.05 (0.04-0.06)	0.07 (0.03-0.08)	0.02 (0.02-0.03)	0.33	0.001
(5BTHF+5ATHF)/THE	1.02 (0.85-1.23)	0.87 (0.78-1.24)	1.11 (0.82-1.28)	0.25	0.99
DHEA/(THE+5BTHF+5ATHF)	0.07 (0.03-0.10)	0.07 (0.03-0.34)	0.18 (0.05-0.31)	0.60	0.14
5BTHF/5ATHF	1.56 (0.98-1.84)	0.93 (0.52-1.46)	0.90 (0.67-1.16)	0.06	0.003
ETIO/ANDRO	0.97 (0.72-1.16)	0.58 (0.45-0.92)	0.87 (0.42-1.15)	0.004	0.33

Table 6.3.1 Urinary steroid metabolites and their ratios in men with HNF1A-MODY vs T2D and healthy controls, *significant p value <0.032 (in bold) corrected for FDR (false discovery rate) of 10% using Benjamini-Hochberg procedure; CTR, controls.

6.3.2 Enzymes with altered activity in HNF1A-MODY

Due to the fact that some crucial enzymes in steroidogenesis convert multiple substrates at different levels of the pathway, the ratios derived from steroid metabolites are regarded more relevant than single steroid measures (Wooding and Auchus 2013). However, individual steroid levels can be analysed to determine if the ratio was driven by a bigger change of a substrate or a product of the enzyme.

Urinary Steroids and their ratios	Median urinary steroid/ ratio (IQR)			p value*	
	HNF1A-MODY	T2D	Control	HNF1A vs T2D	HNF1A vs CTR
Women					
Individual steroids					
Etiocholanolone	650.5 (528.0-2425.0)	680.5 (452.0-915.0)	753.5 (451.0-1095.0)	0.41	0.91
Androsterone	400.5 (268.0-1417.0)	805.5 (389.0-1059.0)	747.0 (508.0-994.0)	0.56	0.34
5-PD	129.0 (54.0-447.0)	191.5 (109.0-269.0)	68.5 (43.0-135.0)	0.84	0.10
Pregnadienol	31.0 (15.0-66.0)	39.5 (23.0-85.0)	31.0 (25.0-57.0)	0.39	0.81
THB	97.5 (48.0-233.0)	59.5 (33.0-89.0)	46.0 (36.0-63.0)	0.09	0.05
TH-DOC	12.4 (4.0-35.0)	8.3 (2.0-27.0)	9.5 (6.0-24.0)	0.42	0.89
PD	165.0 (65.0-409.0)	126.5 (61.0-394.0)	192.0 (109.0-326.0)	0.48	0.74
17-HP	47.5 (18.0-199.0)	22.0 (16.0-60.0)	87.0 (28.0-148.0)	0.34	0.67
THS	63.5 (26.0-116.0)	34.5 (23.0-70.0)	33.0 (22.0-43.0)	0.41	0.04
F	126.0 (27.0-189.0)	66.5 (36.0-102.0)	28.5 (20.0-46.0)	0.54	0.02
6B-OHF	225.0 (60.0-394.0)	98.0 (53.0-146.0)	55.5 (46.0-82.0)	0.12	0.02
5BTHF	1201.5 (382.0-2376.0)	821.5 (420.0-1440.0)	593.0 (415.0-812.0)	0.42	0.11
5ATHF	397.5 (152.0-1121.0)	516.0 (323.0-1185.0)	441.5 (237.0-563.0)	0.54	0.86
E	154.0 (25.0-229.0)	83.5 (64.0-116.0)	45.0 (31.0-57.0)	0.46	0.04
THE	1687.0 (755.0-3511.0)	1294.0 (672.0-2189.0)	1276.5 (894.0-1628.0)	0.48	0.54
Steroid ratios					
F/E	0.77 (0.67-1.06)	0.83 (0.70-0.96)	0.63 (0.52-0.82)	0.91	0.01
(F+E)/(5BTHF+5ATHF+THE)	0.06 (0.04-0.07)	0.06 (0.05-0.07)	0.03 (0.02-0.04)	0.72	0.01
(5BTHF+5ATHF)/THE	0.93 (0.76-1.32)	0.97 (0.86-1.24)	0.80 (0.73-0.93)	0.70	0.10
DHEA/(THE+5BTHF+5ATHF)	0.03 (0.02-0.05)	0.03 (0.01-0.04)	0.07 (0.03-0.19)	0.29	0.05
5BTHF/5ATHF	2.20 (1.81-3.78)	1.29 (1.02-2.17)	1.61 (1.24-2.16)	0.01	0.01
ETIO/ANDRO	1.71 (1.17-2.32)	0.97 (0.74-1.46)	1.17 (0.90-1.49)	0.005	0.02

Table 6.3.2 Urinary steroid metabolites and their ratios in women with HNF1A-MODY vs T2D and healthy controls, *significant p value <0.032 (in bold) corrected for FDR (false discovery rate) of 10% using Benjamini-Hochberg procedure. All abbreviations expanded in the text or Table 6.3.1 above.

6.3.2.1 11 β -hydroxysteroid dehydrogenases

11 β -hydroxysteroid dehydrogenases interchange active cortisol to inactive cortisone, therefore a change in their activity could explain the lean phenotype or low insulin resistance of subjects with HNF1A-MODY.

There was no significant difference in the ratio describing HSD11B1 activity [(5BTHF+5ATHF)/THE] between the groups. However, the activity of HSD11B2, defined by an increased ratio of cortisol/cortisone, was significantly lower in subjects with HNF1A-MODY when compared to healthy controls in both genders (p=0.004 in men and 0.01 in women, Table 6.3.1-2, Figure 6.3.1). A ratio F+E/(5BTHF+5ATHF+THE), also describing a

decreased activity of HSD11B2, was similarly higher in subjects with HNF1A-MODY when compared to the non-diabetic controls in both genders ($p=0.001$ in men and $p=0.06$ in women, Table 6.3.1-2, Figure 6.3.1).

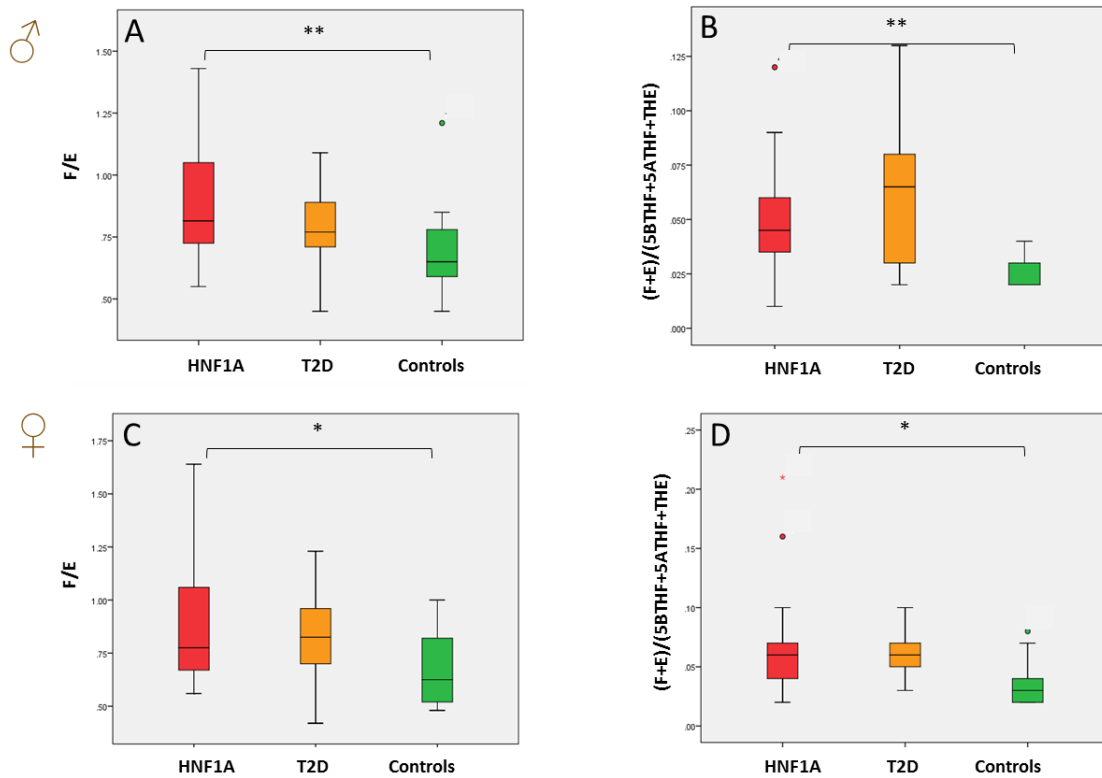


Figure 6.3.1. Ratios describing the activity of 11 β -hydroxysteroid dehydrogenase type 2 (HSD11B2) in men (A-B) and in women (C-D) between HNF1A-MODY, T2D and healthy controls. Each box represents the 25th to 75th percentile. Lines inside the boxes represent the median and whiskers extend to 1.5 times the height of the box or, if no value in that range, to the minimum or maximum value; outliers are marked with a coloured circle and extremes with a coloured asterisk; * $p < 0.02$, ** $p < 0.01$ (Mann-Whitney U test); F, cortisol; E, cortisone; THF, tetrahydrocortisol; THE, tetrahydrocortisone.

Both ratios are consistent with a more significant increase of active cortisol than inactive cortisone in both genders. There was no significant difference in the activity of both 11 β -hydroxysteroid dehydrogenases between subjects with HNF1A-MODY and T2D (Table

6.3.1-2) suggesting that this difference is not HNF1A-MODY specific but related to diabetes in general.

Gender dimorphism of the adrenal steroids is well known and not surprising in the context of steroidogenesis pathways (Figure 1.7.2), where the same precursors could be converted to glucocorticoids or mineralocorticoids as well as sex steroids, which obviously differ between genders.

However, despite the previously reported gender-specific differences in many of the adrenal steroids, there was no difference in the ratios describing the activity of HSD11B2 within each group when the ratios in men were compared to those in women retrospectively (Figure 6.3.2 A).

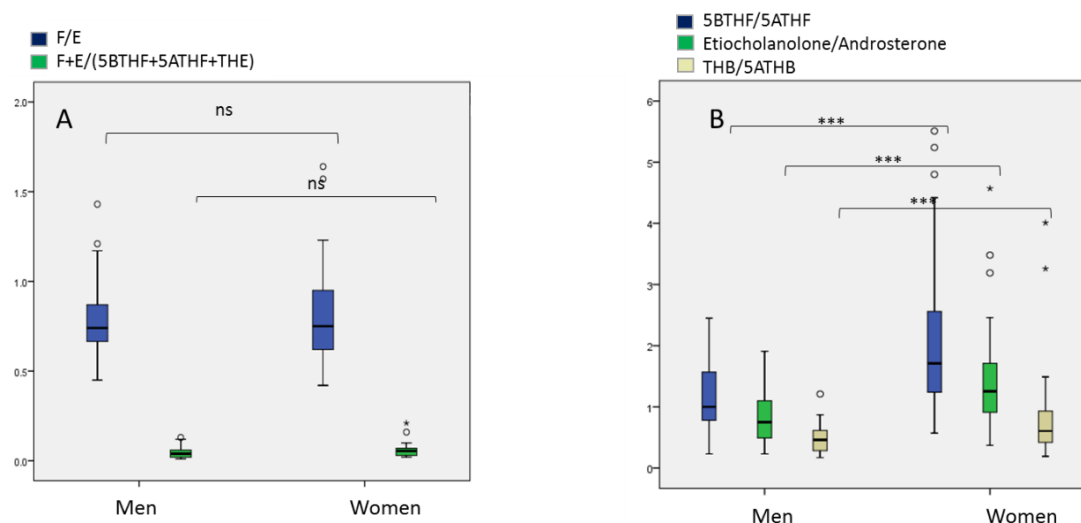


Figure 6.3.2 Ratios describing HSD11B2 activity (A) in men versus women showing no gender difference and ratios defining 5 α - and 5 β -steroid reductases (B) in men and women with a strong gender dimorphism. Each box represents the 25th to 75th percentile. Lines inside the boxes represent the median and whiskers extend to 1.5 times the height of the box or, if no value in that range, to the minimum or maximum value; outliers are marked with a circle and extremes with a small asterisk; ns, non-significant p value; ***, p,0.0001 (Mann-Whitney U test).

In women, there was no statistically significant correlation between ratios defining the activity of HSD11B2 and HbA1c, FPG, duration of diabetes, BMI, BP, creatinine, statin or insulin treatment across all groups examined (all p values non-significant).

In men, the ratios for 11BHSD2 did not correlate with any of the above-mentioned variables, except for the ratio of F+E/(5BTHF+5ATHF+THE) which correlated with FPG ($r=0.364$, $p=0.013$).

This implies that the significant differences in the above ratios were not a result of different diabetes duration or a different insulin resistance in both types of diabetes.

6.3.2.2.5 α - and 5 β -steroid reductase

Similarly to 11 β -HSD type 2, 5 α - and 5 β -steroid reductases deactivate active glucocorticoid, cortisol and thus if they are different in HNF1A-MODY this could provide an explanation as to why those subjects are phenotypically different than individuals with T2D.

Steroid A-ring reductases (5 α - and 5 β -steroid reductase) inactivate cortisol to 5 α -dihydrocortisol (5ADHF) and 5BDHF respectively, which are then converted by 3 α -hydroxysteroid dehydrogenase (HSD3A) to 5 α -tetrahydrocortisol (5ATHF) and 5BTHF and finally to 11 β -hydroxy-androsterone and 11 β -hydroxy-etiocholanolone respectively (Figure 6.2.1). The mRNA expression of *Srd5a1* (the only transcript of this enzyme in mice) was not altered in *Hnf1a* knock-out mice, thus was not expected to be different in my initial hypothesis.

The global activity of SRD5A and SRD5B is described by ratios of 5BTHF/5ATHF and etiocholanolone/androsterone, where increased ratios suggest raised 5 β -steroid reductase or reduced 5 α -steroid reductase activity.

The median ratio of 5BTHF/5ATHF was significantly raised in the HNF1A-MODY group versus healthy controls in both genders [1.56 (0.98-1.84) vs 0.90 (0.67-1.16), $p=0.003$ in men and 2.20 (1.81-3.78) vs 1.61 (1.24-2.16), $p=0.012$ in women]. The ratio of etiocholanolone/androsterone was only increased in female participants [1.71 (1.17-2.32) vs 1.17 (0.90-1.49), $p=0.019$, Figure 6.3.3].

The difference in the ratio of etiocholanolone/androsterone was even more prominent when HNF1A-MODY was compared to T2D group in both genders ($p=0.004$ in men and $p=0.005$ in women), but 5BTHF/5ATHF remained significantly elevated only in women ($p=0.014$, Table 6.3.1-2, Figure 6.3.3).

To answer which steroid reductase activity is contributing to the above increase in ratios, the individual metabolites were compared. I found that 5 β -metabolites (5BTHF, THE and etiocholanolone) were all significantly higher in HNF1A-MODY patients when compared to healthy controls in men ($p=0.004$, 0.017 and 0.017, Figure 6.3.4), while 5 α -metabolites (5ATHF and androsterone) were similar in all groups.

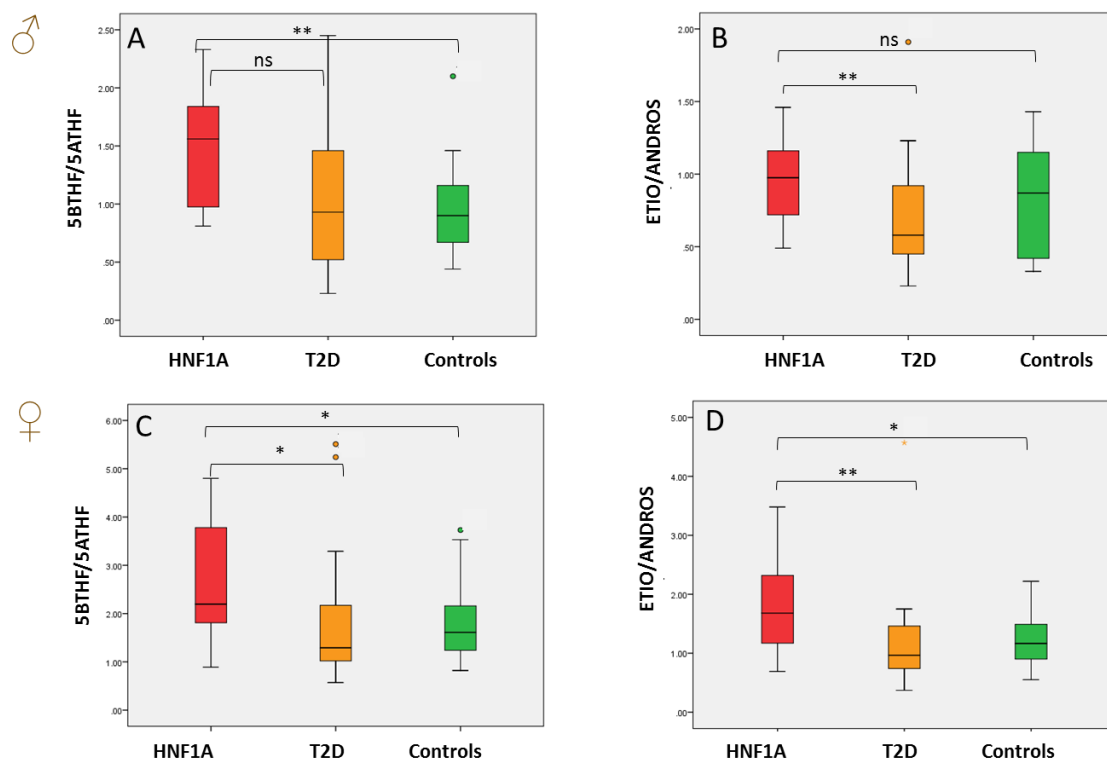


Figure 6.3.3 Ratios describing the activity of total 5 α -steroid reductase SRD5A in men (A-B) and in women (C-D) between HNF1A-MODY, T2D and healthy controls. Each box represents the 25th to 75th percentile. Lines inside the boxes represent the median and whiskers extend to 1.5 times the height of the box or, if no value in that range, to the minimum or maximum value; outliers are marked with a coloured circle; * $p < 0.02$, ** $p < 0.01$; ns, non-significant p value (Mann-Whitney U test).

There was no statistically significant difference in the individual 5 β - or 5 α -steroid metabolites in women. The trend was similar to the one in men for 5 β THF but did not reach statistical significance ($p = 0.1$). Thus, I concluded that the increased steroid reductase ratios in men were mainly due to the increased 5 β -steroid reductase activity.

Both ratios defining the activity of SRD5A and SRD5B were similar between T2D and healthy control groups in both genders.

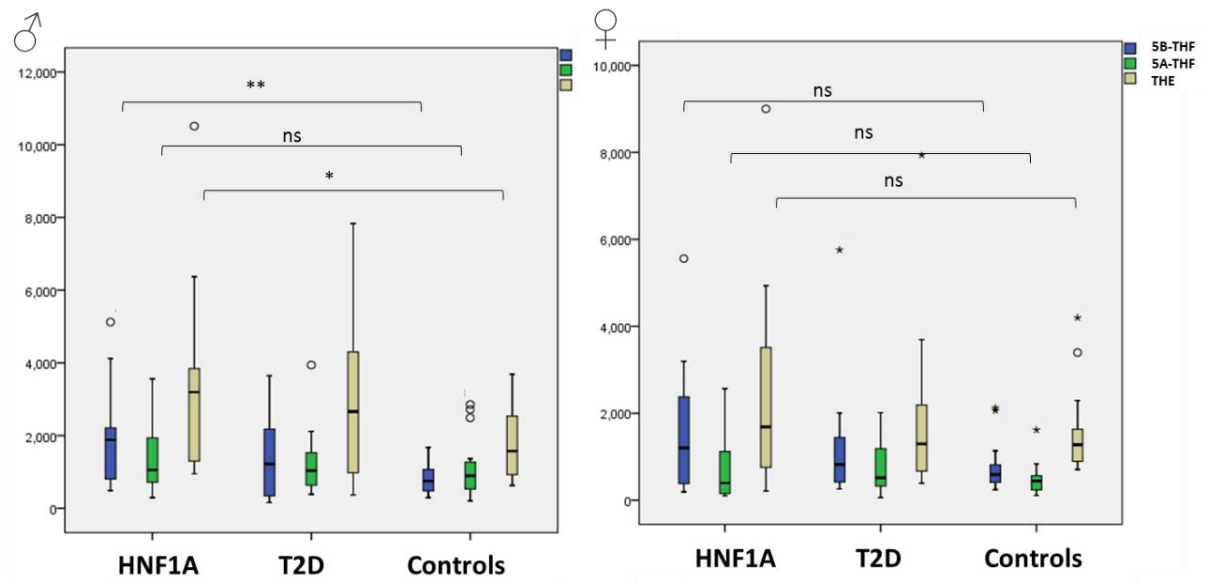


Figure 6.3.4 5 α - (5ATHF) and 5 β -cortisol metabolites (5BTHF, THE) plotted together in men (σ) and women (ρ). 5 β -metabolites higher in men with HNF1A-MODY vs healthy controls, while no statistically significant difference in 5 α -metabolites between those groups. No difference in women. Each box represents the 25th to 75th percentile. Lines inside the boxes represent the median and whiskers extend to 1.5 times the height of the box or, if no value in that range, to the minimum or maximum value; outliers are marked with a circle and extremes with a small asterisk; * $p < 0.05$. ** $p < 0.01$.

Generally, men had higher 5 α - and 5 β -steroid metabolites than women in all examined groups, which is consistent with gender dimorphism of those steroid metabolites reported in the literature previously. There was also a gender-related difference in the ratios describing the activity of SRD5A and SRD5B, which were higher in women than in men in all groups ($p = 1 \times 10^{-3}$ - 6×10^{-6} , Figure 6.3.2B, Table 6.3.1-2). This is also consistent with previous literature (Finken et al. 1999).

In women, there was no significant correlation between ratios defining the activity of SRD5B and HbA1c, FPG, duration of diabetes, BMI, BP, statin or insulin treatment across all groups (all p values non-significant).

In men, the ratio of Etiocholanolone/Androsterone positively correlated with diabetes duration ($r=0.46$, $p=0.007$) and creatinine ($r=0.52$, $p=0.004$). A correction of the ratio of Etiocholanolone/Androsterone for the duration of diabetes retained a significant p value of 0.014. The ratio of 5BTHF/5ATHF correlated with FPG ($r=0.47$, $p=0.011$) which did not differ between diabetes groups and therefore not affecting the results. There was no correlation of the above ratios with statin or insulin treatment in either gender.

6.3.2.3 Other altered steroidogenesis enzyme activities

The ratio used to define the activity of HSD3B and hypothesised initially to be reduced in HNF1A-MODY patients when compared to non-diabetic controls [DHEA/(THF+5ATHF+THE)] was not significantly different between examined groups in both genders.

6.3.2.4 Correlation of uncorrected versus creatinine-corrected individual steroid metabolites

Renal function, measured by eGFR (estimated glomerular filtration rate), affects the concentration of substances excreted in urine. Therefore, I excluded subjects with plasma creatinine $>160\mu\text{mol/L}$ and the combined mean creatinine of all groups was $79.5 (\pm 24.4)\mu\text{mol/L}$ (Table 6.2.1). The values of all urinary steroid metabolites were corrected for

urinary creatinine using an equation described in methods (section 6.2.3.2) for the first 50 subjects (HNF1A-MODY and T2D).

The Pearson correlation coefficient was used for all individual uncorrected steroid metabolites versus creatinine-corrected values. All individual steroids correlated significantly with their creatinine-corrected counterparts (Pearson's $r=0.73-0.98$, $p<0.0001$ for all) and Figure 6.3.5 shows correlation scatterplots for six steroid examples used most commonly in the discussed ratios (androsterone, etiocholanolone, cortisol, 5BTHF, cortisone and THE).

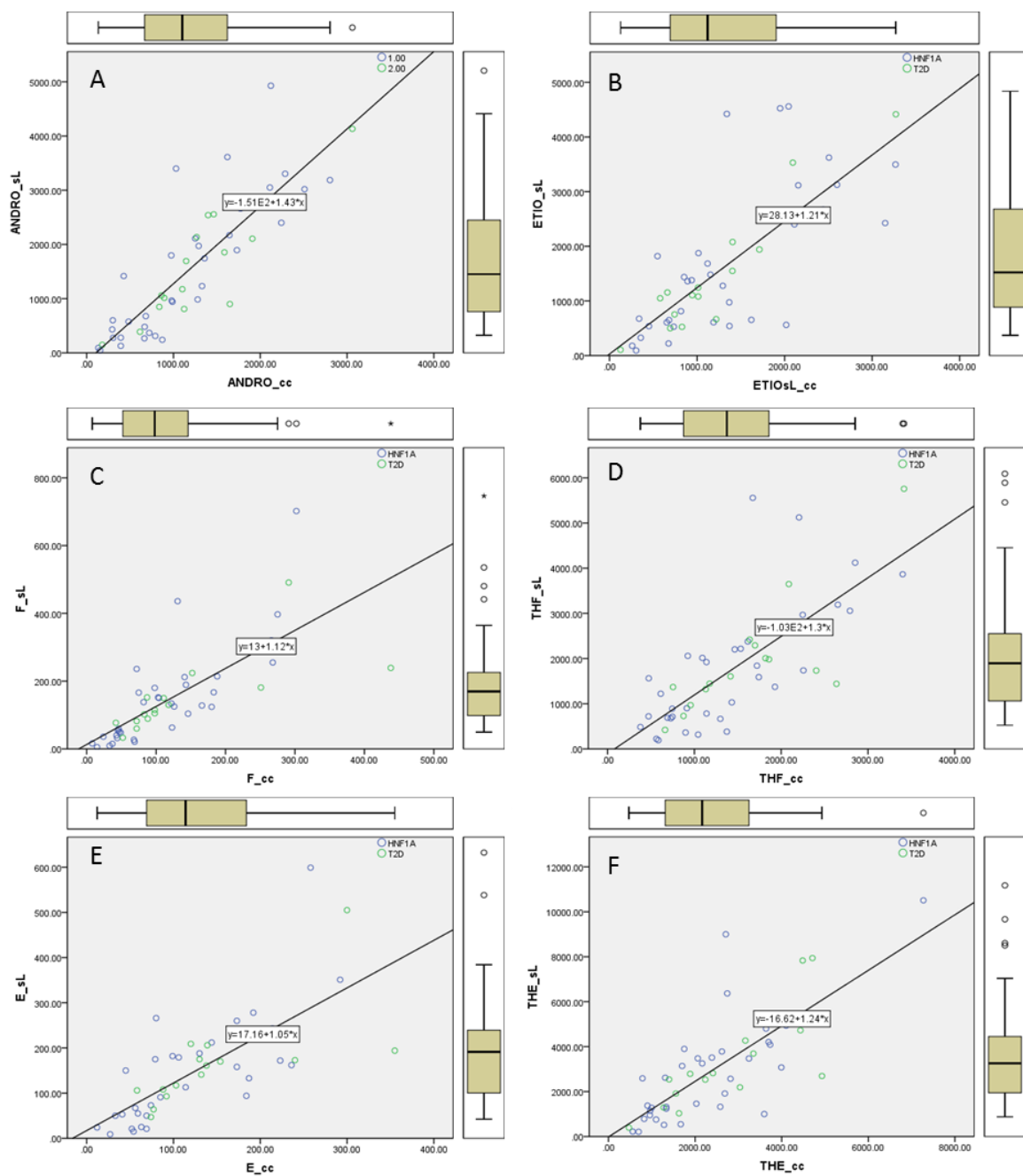


Figure 6.3.5 Pearson correlation scatter plots for androsterone (A), Etiocholanolone (B), cortisol (C), THF, tetrahydrocortisone (D), cortisone (E) and THE, tetrahydrocortisone (F) comparing uncorrected (sL, y axis) vs creatinine corrected (cc, x axis) steroid levels.

6.3.2.5 Urinary steroid ratios as biomarkers of HNF1A-MODY

Two ratios describing the activity of 5 β -steroid reductase were significantly higher in subjects with HNF1A-MODY when compared to the individuals with T2D. The ratio of etiocholanolone/androsterone was higher in HNF1A-MODY in both men and women, while 5BTHF/5ATHF was only significantly higher in women. Thus, both were considered as potential biomarkers of HNF1A-MODY when differentiating this type of diabetes from T2D. I performed a receiver operating characteristic (ROC) curve analysis to assess their classification performance (Figure 6.3.6).

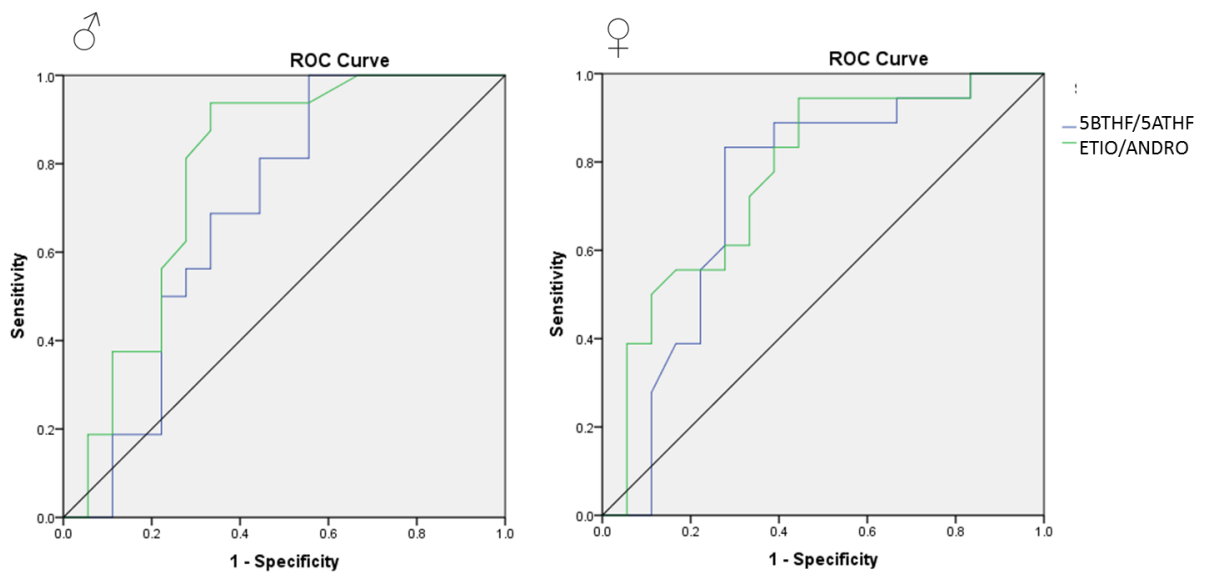


Figure 6.3.6 Receiver operating characteristic (ROC) curves depicting performance of both ratios defining the activity of steroid reductases by ratios of 5 β -metabolites per 5 α -metabolites of cortisol as a screening tool for HNF1A-MODY; 5BTHF/5ATHF in blue and ETIO/ANDRO in green; THF, tetrahydrocortisol; ETIO, etiocholanolone; ANDRO, androsterone. ROC curve of both ratios in men on the left, in women on the right.

The ratio of etiocholanolone/androsterone achieved a C-statistic of 0.78 (95% CI 0.62-0.94) in men and 0.77 (0.61-0.93) in women. This translated to a sensitivity of 81% and

specificity of 72% for the cut-off 0.72 in men and a sensitivity of 78% and specificity of 67% for the cut-off 1.16 in women. Based on this, neither of these would be considered a good biomarker candidate for HNF1A-MODY in both genders, however discriminatory performance of the ratio of etiocholanolone/androsterone was satisfactory.

The ratio of 5BTHF/5ATHF had C-statistics of 0.69 (95% CI 0.51-0.88) in men and 0.74 (0.56-0.91) in women and would not support its use as a biomarker of HNF1A-MODY (Figure 6.3.6).

6.4 Discussion

I have described the first urinary steroid profile in individuals with HNF1A-MODY. The activity of two enzymes deactivating cortisol, 11 β -hydroxysteroid dehydrogenase type 2 and 5 β -steroid reductase were significantly altered in subjects with HNF1A-MODY when compared to age- and BMI-matched healthy controls.

I found that the activity of 5 β -steroid reductase was significantly increased in subjects with HNF1A-MODY, distinguishing them from both T2D and healthy controls. The individual 5 β -metabolites: 5BTHF, etiocholanolone and THE, but not 5 α -metabolites, were significantly increased in men with HNF1A-MODY.

SRD5B (AKR1D1) provides an alternative pathway to deactivate cortisol in the liver and is an important enzyme for the secondary bile acid synthesis pathway (Chen and Penning 2014). Both pathways have the potential to regulate the metabolic phenotype.

An elevation of SRD5B activity in individuals with HNF1A-MODY is also consistent with the previous report that bile acid from the classic bile acid synthesis pathway, 7 α -hydroxy-4-cholesten-3-one, was increased in HNF1A-MODY participants (n=19) when compared to age- and BMI-matched healthy controls (n=15) (Ekholm et al. 2013).

The upregulation of *Cyp7a1* and downregulation of *Cyp7b1* (both present in the primary bile acid synthesis pathway) were also reported in the hepatocytes from the *Hnf1a* knock-out mice (Servitja, Pignatelli et al. 2009). An increase in CYP7A1 results in an increase of its product, 7 α -hydroxy-4-cholesten-3-one, which would result in an increase in SRD5B activity (downstream from CYP7A1 in secondary bile acid synthesis) (Reihner et al. 1989) and is in line with my findings.

Ekholm *et al.* also performed the functional characterisation of 5 *HNF1A* variants (p.L107I, p.R131Q, p.T260M, p.P291fs, p.P379A) from subjects with HNF1A-MODY using the human hepatoma cell line (HuH7) and a bile acid nuclear receptor (Farnesoid X factor, *FXR*) promoter (Ekholm, Nilsson et al. 2013). In the human liver, the activation of FXR leads indirectly to the repression of CYP7A1 via the upregulation of small heterodimeric partner (SHP). Variants p.L107L, p.T260M, p.P291fs and p.R131Q all reduced the transcription activity of the FXR promoter in HuH7 cells to less than 40% of the wild-type HNF1A (p<0.001), which would be expected to alleviate the repression of *CYP7A1* and upregulate *SRD5B*. Variant p.P379A did not affect the transcription of FXR. The above experiments supported an indirect effect of *HNF1A* variants via FRX on bile acid production and an increase of the product of CYP7A1 in HNF1A-MODY individuals with the four variants

performing similarly. Again, variant p.P379A behaved similarly to the WT HNF1A in those experiments.

The apical sodium-dependant bile salt transporter (*SLC10A2*) facilitates reabsorption of bile acids in the terminal ileum and was also found to be downregulated in *Hnf1a* ^{-/-} mice (Servitja, Pignatelli et al. 2009). An experiment with the above *HNF1A* variants and the *SLC10A2* promoter using a human intestinal cell line (Caco2) showed decreased transcription activity of the *SLC10A2* promoter by p.L107L, p.T260M, p.P291fs and p.R131Q (p <0.001) and p.P379A performed again similarly to the WT HNF1A (Ekholm, Nilsson et al. 2013). This experiment suggests an altered bile acid reabsorption in HNF1A-MODY individuals.

The change in bile acids has implications on the composition of the gut microbiota which was also reported as different in HNF1A-MODY subjects versus healthy controls and individuals with T2D (Mrozinska et al. 2016). Participants with HNF1A-MODY (n=10) had decreased *Bacteroides* when compared to BMI-matched healthy controls (n=16) and increased *Turicibacterales* when compared to the individuals with T2D (n=23).

My data also showed that there is no difference in the global activity of HSD11B1 based on the urinary steroid profile in subjects with HNF1A-MODY or T2D when compared with healthy controls. This data in subjects with T2D is consistent with previous literature (Valsamakis, Anwar et al. 2004, Tomlinson, Finney et al. 2008).

However, I have detected that the activity of 11 β -hydroxysteroid dehydrogenase type 2 was significantly reduced in participants with HNF1A-MODY and T2D when compared to healthy controls in both genders. A decreased HSD11B2 activity would be expected to increase renal cortisol levels, which may drive an increase in total cortisol metabolite production. The local impairment of inactivation of cortisol is expected to elevate blood pressure (BP) due to the increased binding of cortisol to the mineralocorticoid receptor (MR) in the renal collecting ducts.

In my study, only men with HNF1A-MODY had significantly higher systolic BP (including 2 individuals on treatment) when compared to age- and BMI-matched healthy controls [139 (130.5-148.5) mmHg vs 125 (117-128) mmHg respectively, $p=0.02$]. There was no difference in BP between subjects with HNF1A-MODY and T2D in both genders, consistent with the fact that individuals with T2D expressed a similar reduction in the activity of HSD11B2 but they were more often on anti-hypertensive treatment (7 men and 12 women).

The decreased activity of HSD11B2 is inconsistent with previous reports which found no difference in the urinary F/E ratio in obese men with T2D, although the serum F/E ratio was increased. Circulating levels of cortisol or cortisone mainly reflect the activity of hepatic 11 β -HSDs (Valsamakis, Anwar et al. 2004). The study comparing BMI-matched obese individuals with impaired glucose tolerance (IGT) versus healthy participants also showed no significant difference in the urinary F/E ratio between groups (Tomlinson, Finney et al. 2008). In both above-mentioned studies, participants had a significantly

higher BMI (mean 32-34 kg/m²), while in my study subjects were slimmer (mean BMI 26.5 kg/m²). However, neither cortisol, cortisone nor F/E ratio correlated with BMI in my study.

Subjects with HNF1A-MODY and T2D had a statistically significant increase in urinary free cortisol levels in both genders when compared to non-diabetic controls. A reduced 11BHSD2 activity increased levels and the half-life of renal cortisol which feeds back to the hypothalamus-pituitary-adrenal (HPA) axis so that circulating cortisol levels are not changed. Even in individuals with apparent mineralocorticoid excess syndrome due to genetic defects in *HSD11B2*, the circulating cortisol levels are normal as the HPA axis resets (Palermo et al. 2004) and the change of HSD11B2 activity is at a tissue-specific level, in this case the kidney. The changes in 11BHSD2 could be important in helping to understand BP regulation and salt/water balance in subjects with HNF1A-MODY and T2D, and in recognising that specific therapies such as MR antagonists may be more appropriate.

It was previously reported that up to 30% of individuals with HNF1A-MODY have associated liver adenomas, often with biallelic inactivation of *HNF1A* in the adenoma cells (Bluteau et al. 2002). The co-segregation of disease-causing *HNF1A* variants and liver adenomatosis was previously reported (Reznik et al. 2004), thus its role as a tumour-suppressor gene was hypothesised. It is well established that hepatocellular adenomas occur in women taking oral contraceptives (Rooks et al. 1979) or in men using anabolic steroids (Socas et al. 2005), hence why *HNF1A* germline mutations and elevated steroids

could contribute to liver tumorigenesis. This hypothesis would require further work to prove.

The limitation of my study was measuring urinary steroids in morning spot urine samples and then calculating their concentration per litre of urine rather than using 24-hour urine collection. However, the comparison of urinary steroid results measured in mid-morning spot urine samples versus 24-hour urine collection was performed by Nanus *et al.*, who found a high correlation of both method results, especially for steroid metabolite ratios (Nanus et al. 2015). I have also shown that it is not necessary to measure urinary creatinine and correct urinary steroids for creatinine, as uncorrected urinary steroid values correlated very well with creatinine-corrected ones. This is true for plasma creatinine levels less than 160 μ mol/L, as measuring any urinary substances as proxy for blood levels above this value is unreliable.

6.5 Conclusions

I have described the first urinary steroid profile in HNF1A-MODY individuals. The activity of two enzymes deactivating cortisol, 11 β -hydroxysteroid dehydrogenase type 2 and 5 β -steroid reductase were significantly altered in subjects with HNF1A-MODY when compared to healthy controls. The function of SRD5B was also significantly higher in subjects with HNF1A-MODY than in participants with type 2 diabetes. My findings suggest that in subjects with HNF1A-MODY, an increased activity of SRD5B may have implications on their phenotype by affecting bile acid composition and gut microbiota. This gives background for further research.

Chapter 7

Overall discussion

In my thesis:

- I have described 26 rare non-synonymous *HNF1A* variants found from sequencing of 30 probands and 8 relatives with non-autoimmune diabetes diagnosed below 45 years of age. I have shown the complexity of assigning a functional effect to *HNF1A* alleles and concluded that 50% of those alleles are functionally deleterious and are likely to contribute to the diabetic phenotype of the carriers, as discussed in chapter 3.
- I have also estimated the percentage of individuals carrying functionally deleterious *HNF1A* alleles in Croatia as 75 cases per million in the general population, which has not been reported previously.
- In chapter 4, I have illustrated that functional studies complement the interpretation of the effect of *HNF1A* alleles and, when protein function is significantly impaired, provide strong evidence for the allele to be contributing to the diabetes status.
- In chapter 5, I discussed biomarkers for identifying those with damaging *HNF1A* alleles of which CRP, being inexpensive and widely available, is already used in Oxford MODY clinic and in the genetic laboratory in Exeter. Glycan GP30 discriminated individuals with damaging *HNF1A* alleles from those with young-onset type 2 diabetes and no *HNF1A* variants better than CRP, but currently measuring the N-glycan profile is too expensive and lacks high throughput assay.

- The urinary steroid profile described in chapter 6 did not identify a new biomarker of young adult onset non-autoimmune diabetes in individuals carrying functionally deleterious *HNF1A* alleles, but provided some insights into the pathophysiology of subjects with this type of diabetes, giving basis for future research.

Based on sequencing *HNF1A* from 1032 individuals with C-peptide positive and GAD antibody negative diabetes diagnosed before 45 years of age, I have estimated the percentage of individuals carrying functionally deleterious *HNF1A* alleles which are likely to contribute to the diabetic phenotype of the carriers, as 1.89% in Croatia and 1.94% in the UK in the above group. Not all of those individuals have HNF1A-MODY phenotype so the prevalence of this type of monogenic diabetes is likely to be lower than percentage of individuals with functionally deleterious *HNF1A* alleles. This prevalence is lower than the prevalence of HNF1A-MODY reported previously in the sample of individuals with non-autoimmune diabetes diagnosed before 45 years (4.6%) (Thanabalasingham, Pal et al. 2012). The difference is related to the fact that some *HNF1A* alleles were previously reported as MODY-causing (such as p.G31D or p.G606S) but subsequent data, including large-scale sequencing projects such as GnomAD, did not support it. Another reason for overestimating the prevalence of HNF1A-MODY would be not taking into account the functional spectrum of *HNF1A* alleles which could also contribute to the T2D phenotype. Therefore, I think it is more appropriate to avoid using the term “HNF1A-MODY” in the study discussed in my thesis and use “young adult onset non-autoimmune diabetes in individuals carrying functionally deleterious *HNF1A* alleles” when describing the prevalence or biomarkers.

An increase in large-scale sequencing projects and the introduction of next-generation sequencing into research and clinical practice have revealed a large number of genetic variants. The interpretation of these variants remains a challenge in the field of any genetic disorder. Hence, I systematically assessed all available evidence to assign the functional effect of rare *HNF1A* alleles found in the study presented in this thesis. This unveiled the complexity of the genotype-phenotype relationship and continuity of the allelic functional spectrum.

The consequences of this continuum had been reported as a variety of phenotypes in the many MODY pedigrees since the discovery of MODY genes (Lehto et al. 1997, Kavvoura et al. 2014, Knebel et al. 2016). However, the plausible explanation of variable allelic effect had initially been overlooked or ignored in the context of traditional classification of non-autoimmune diabetes. The demonstration that rare missense coding variants in MODY genes could also be found in subjects with the type 2 diabetes phenotype or in middle-age non-diabetic individuals blurred further genetic boundaries between MODY, T2D and non-diabetic status (Flannick, Beer et al. 2013, Lek, Karczewski et al. 2016). This again supported the continuum of the allelic effect. This has significant clinical repercussions on diagnosing rare forms of monogenic diabetes, classification of diabetes and an estimation of the true prevalence of MODY. As boundaries between MODY and T2D represent a grey zone, it is not clear which cases should be included when the prevalence of MODY is calculated.

It should be noted that there are alleles with high penetrance such as p.P291fs which can cause a young adult-onset diabetes in majority of affected family members such as “family C” reported by Glucksmann *et al.* (Glucksmann, Lehto et al. 1997). In this family, all 4 children and all 9 grandchildren had diabetes onset at 8-26 years with the exception of one member diagnosed at 48 years.

One of the challenges discussed in the introduction of my thesis was a low diagnostic rate of MODY and that around 80% of individuals with MODY are misdiagnosed as more common types of diabetes. The classic criteria for genetic screening for MODY miss individuals who present above 25 years of age or do not have a family history of diabetes (50% of cases with HNF1B-MODY are due to *de novo* variants). An online MODY calculator and its smartphone App version are available and if widely advertised could increase the number of subjects diagnosed with the typical MODY phenotype. However, the MODY calculator does not make it possible to calculate the probability of monogenic diabetes for subjects with diabetes onset above 35 years: it is weighted towards those with a “classic” MODY presentation that were used to define the model (Shields, McDonald et al. 2012). Taking into account the boundaries between MODY and T2D which are too blurred to be separable, especially given the prior odds strongly in favour of T2D, it is hard to improve upon the performance of this calculator in older age of diabetes onset.

The extension of the age of diabetes onset up to 45 years in the inclusion criteria for genetic testing used in our study would inevitably include more subjects with type 2 diabetes. This raises the question: which type of diabetes do the individuals with a

borderline phenotype (diabetes onset at 35-45 years, BMI above 25-30kg/m² and some features of metabolic syndrome) have? Are they really missed cases of MODY or perhaps subjects with T2D?

The spectrum of the functional effect of *HNF1A* alleles correlates with but does not map directly onto the phenotypical continuum of non-autoimmune diabetes. This means that an allele from the middle of the spectrum can also cause the HNF1A-MODY phenotype in an individual with diabetes but less often than for example p.P291fs or p.T260M (both with severely impaired protein function). On the other hand, showing experimentally that a particular *HNF1A* allele could potentially cause the MODY phenotype does not necessarily mean that it would cause it in a particular individual with diabetes. Due to a spectrum of the allelic functional effect and other factors influencing the final phenotype of an individual, people carrying *HNF1A* variant alleles show a variation of the clinical phenotype, even within the same family (Bellanne-Chantelot, Levy et al. 2011, Fajans and Bell 2011).

The phenotypical spectrum of non-autoimmune diabetes is consistent with the “palette” model of diabetes which illustrates each diabetic individual’s phenotype as a consequence of interactions between multiple genetic, pathophysiological (e.g. β -cell function, insulin resistance, incretin effect) and environmental factors such as lifestyle or diet (McCarthy 2017). The influence of these multiple factors results in different clinical phenotypes from MODY to typical T2D or an increased risk of T2D in non-diabetic individuals. This model will also fit the progression of diabetes over time, when an individual with any type of

diabetes moves along the phenotypical continuum after gaining or losing weight or with the progression of β -cell dysfunction, which also applies to HNF1A-MODY.

Variant p.P291T in *HNF1A* is an example of an allele with a modest decrement in function which sits in the middle of the allelic spectrum and contributes to a phenotype closer to T2D. This allele had been previously reported as MODY-causing (Bellanne-Chantelot, Carette et al. 2008, Thanabalasingham, Shah et al. 2011). However, p.P291T is predicted by all three *in silico* tools to be neutral to protein function. In the study presented in this thesis, all three British family members carrying this *HNF1A* allele were obese (BMI 31-38.8kg/m²) with diabetes diagnosed at 27, 33 and 63 years of age. The proband and his cousin were not sensitive to sulphonylureas and quickly progressed to other treatments. This allele was present in 3 out of 4 tested subjects with diabetes in this family (non-diabetic individuals were not tested). The experimental work discussed in chapter 4 showed a moderate reduction of transactivation (59% of WT) but normal protein expression and DNA binding. Hence, it did not strongly support the damaging effect of this *HNF1A* allele. Another individual from our MODY clinic with the same *HNF1A* allele was diagnosed with diabetes at age of 15 years and was slim at diagnosis. Therefore, this allele is probably increasing risk of T2D in the first proband and likely contributing to the MODY phenotype in the second proband.

Another *HNF1A* allele with a less severe phenotypical impact than previously assumed (Colclough, Bellanne-Chantelot et al. 2013) is p.G31D, which was initially considered as MODY-causing (Boutin et al. 1997, Guazzarotti et al. 1999). Subsequently, it was noted to

lack co-segregation with young adult onset diabetes in families (Ellard, Colclough personal communication) (Juszczak A 2014) and to have a minor allele frequency of 0.1% in non-Finnish European individuals in GnomAD, which is more common than expected for HNF1A-MODY-causing alleles (Lek, Karczewski et al. 2016). This variant is localized in the dimerization domain and previous functional characterisation of this allele showed normal dimerization and DNA binding of this protein variant (Rose et al. 2000). The patient from our MODY clinic carrying this *HNF1A* allele presented with diabetes at the age of 21 years and was initially given a diagnosis of T1D. Ten years later, she was found to have an upper normal level of C-peptide and referred for genetic screening for MODY which confirmed the presence of *HNF1A* variant allele p.G31D and her diagnosis was changed to HNF1A-MODY. However, she was not sensitive to gliclazide, and despite a maximal gliclazide dose and a high dose of insulin, good control of hyperglycaemia was not maintained. At that stage, features of severe metabolic syndrome were noted and linked to diagnosis of secondary diabetes due to total body irradiation in childhood for treatment of acute leukaemia (Juszczak A 2014). This case also reflects the importance of re-evaluation of the initial assignment of the allelic functional effect and diagnosis when more evidence becomes available.

Next-generation sequencing methods are increasingly used in clinical settings including the diagnosis of MODY. However, in the clinical settings, the assessment of their results depends on clinical selection bias as they are often performed on high-risk individuals (Goring et al. 2001, Ellard et al. 2013). In our study, we had a relatively unselective

approach to recruitment, excluding only individuals with autoimmune diabetes (T1D and LADA). This allowed a more accurate assessment of the use of hsCRP and glycan profile for selecting those at high risk of HNF1A-MODY. This approach also enabled an illustration of the phenotypical spectrum of *HNF1A* alleles seen in young adults. Dividing subjects with rare *HNF1A* alleles into groups depending on the available bioinformatic, genetic, clinical and functional characterisation data, discussed in chapter 3, was to some extent limited by the disagreement of different assessment components and by incomplete co-segregation studies. Another limiting factor was the lack of data on response to sulphonylurea (SU) treatment.

It is important to identify *HNF1A* allele carriers who are sensitive to SU, as those subjects would achieve and likely maintain a target HbA1c for many years on that treatment. When the *HNF1A* variant is found in an individual with C-peptide positive and GAD antibody negative diabetes in our MODY clinic, the response to SU is assessed as a first line treatment or added to the current diabetes treatment. However, it is unknown if clinicians should also use SU in mono- or combination therapy in individuals with T2D carrying missense *HNF1A* alleles. This question could be answered by future clinical research.

In chapter 5, I showed that CRP and 3-antenary fucosylated glycans such as GP30 could be used for the selection of individuals for *HNF1A* sequencing. However, recent wider use of next-generation sequencing (NGS) allows sequencing of the whole exome/genome efficiently so Sanger sequencing of the *HNF1A* alone will be used less frequently. Sanger

sequencing is still a method of choice for sequencing one exon looking for a particular variant in the relative of the proband, while NGS is more appropriate in subjects with a typical phenotype of MODY but with no features suggestive of a particular subtype of MODY or in gene discovery studies. Therefore, CRP or GP30 may be more useful as an adjunct to the assessment of the functional effect of *HNF1A* alleles or in the selection of subjects for genetic testing in the subgroup diagnosed with diabetes above 35 years of age when tools such as the MODY probability calculator cannot guide the clinician. The present study could be extended with collaboration of other diabetes centres nationally or internationally as future research. The future work could also evaluate the health economics aspect of using biomarkers for identifying those with damaging *HNF1A* alleles and Sanger sequencing versus use of the MODY calculator and next-generation sequencing in subjects with non-autoimmune diabetes. Naylor *et al.* had used a simulation model to compare costs of health care for subjects with HNF1A-, HNF4A- and GCK-MODY with a genetic diagnosis made at the diagnosis of diabetes versus natural history of the disease and no genetic testing. Their model suggested that screening for MODY has been cost-effective in the USA but it simulated the outcomes of the three most common types of MODY together (Naylor *et al.* 2014). Including patients with GCK-MODY who do not require treatment would reduce total costs of health care if analysed together with other types of monogenic diabetes.

The decisions I made about the functional impact of *HNF1A* alleles are based on a synthesis of a variety of data but the clinical information on the subjects from this study

was excluded. This exclusion allowed to carry out an unbiased assessment of the value of functional classification to make clinical decisions and to avoid circularity. The clinical phenotype may vary between subjects with the same *HNF1A* allele and a functional assessment approach leaving clinical features of a particular individual with diabetes to the end of assessment would be a better approach. When the functional assessment of the allele is completed, it should be available to the clinician who can make a diagnostic decision in the context of the individual patient's phenotype.

Hence, an initial functional classification based on purely experimental data would have the advantage of not incorporating any clinical information *a priori*. An attempt to provide a better interpretation of the functional studies was performed by my colleague Sara Althari from Prof. Gloyn and McCarthy's group in the Oxford Centre for Diabetes, Endocrinology and Metabolism. She performed a principal component analysis using results from three different functional assessments (transcription activity, DNA binding and protein expression by western blotting) of 27 rare non-synonymous *HNF1A* variants (with mine and another colleague from Prof. Gloyn's and McCarthy's group, Dr Bennett's functional data together) to see how the variants cluster based on functional studies only. This approach to the functional assessment could be used to provide a more accurate functional characterisation outcome to contribute to clinical diagnosis.

Taking the functional assays out into a completely undefined set of samples would be the way to overcome the above-mentioned circularity. High-throughput variant screening to

functionally interrogate all possible single-nucleotide substitutions seems an attractive and effective way of assigning a functional effect to all missense variants in *HNF1A*.

The functional interrogation of all possible single-nucleotide substitutions in *HNF1A* (11,989 alleles) using cell-sorting techniques are also being investigated by Althari *et al.* (American Society of Human Genetics, ASHG 2017 abstract). The human liver cell line, HepG2 has been transduced with 226 *HNF1A* variants so far and the cell surface-expressed targets of *HNF1A* were tested. The fluorescently-labelled target *TM4SF4* (Transmembrane 4 Superfamily Member 4 gene, encoding a cell surface glycoprotein that can regulate cell proliferation) achieved the best allelic segregation in flow cytometry and so far, 2-8% of *HNF1A* alleles displayed a loss-of-function effect. This work is on-going and aims to provide an unbiased experimental assessment of the functional effect of all possible missense *HNF1A* alleles.

Multiplex assays of variant effect (MAVEs) such as the one described above provide the advantage of assessing the functional consequences of all possible missense variants for multiple molecular phenotypes. As in large-scale sequencing projects such as GnomAD, the functional data could also be combined to provide a large comprehensive functional database of genetic variation (Starita *et al.* 2017). However, MAVEs would have potential limitations related to all *in vitro* studies as they are mainly cell-based assays that cannot assess all effects of the variant or complex interactions between physiological pathways present in humans.

In conclusion, my thesis has shown the complexity of the interpretation of genetic variation which is crucial for any further research such as the assessment of a potential biomarker for genetic disease, or the diagnosis and management of diabetes in a particular individual with a genetic variant. A functional spectrum of *HNF1A* alleles correlates with the clinical phenotypes of a sample of individuals with diabetes but varies between subjects, even those with the same *HNF1A* variant. Therefore, it must be acknowledged that other genetic and environmental factors also contribute to the final phenotype in that individual. The CRP and glycan GP30 could further complement the assessment of the functional effect of *HNF1A* alleles but these biomarkers have their own limitations. Future research on high-throughput functional assays of variant effect could help form a more reliable interpretation of experimental effects of genetic variants but must be combined with clinical data to provide a more reliable functional and clinical phenotype database which is currently lacking.

References

- Adalat, S., et al. (2009). "HNF1B mutations associate with hypomagnesemia and renal magnesium wasting." J Am Soc Nephrol **20**(5): 1123-1131.
- Adzhubei, I. A., et al. (2010). "A method and server for predicting damaging missense mutations." Nat Methods **7**(4): 248-249.
- Ahmed, A., et al. (2012). "A switch in hepatic cortisol metabolism across the spectrum of non alcoholic fatty liver disease." PLoS One **7**(2): e29531.
- Aken, B. L., et al. (2016). "The Ensembl gene annotation system." Database (Oxford) **2016**.
- Alberts, P., et al. (2002). "Selective inhibition of 11beta-hydroxysteroid dehydrogenase type 1 decreases blood glucose concentrations in hyperglycaemic mice." Diabetologia **45**(11): 1528-1532.
- Alberts, P., et al. (2003). "Selective inhibition of 11 beta-hydroxysteroid dehydrogenase type 1 improves hepatic insulin sensitivity in hyperglycemic mice strains." Endocrinology **144**(11): 4755-4762.
- Arlt, W., et al. (2011). "Urine steroid metabolomics as a biomarker tool for detecting malignancy in adrenal tumors." J Clin Endocrinol Metab **96**(12): 3775-3784.
- Awata, T., et al. (1998). "A novel missense mutation in the homeodomain of the hepatocyte nuclear factor-1alpha/maturity-onset diabetes of the young 3 in a Japanese early-onset type 2 diabetic patient and time-course of glucose-stimulated insulin secretion." Diabetes Care **21**(9): 1569-1571.
- Bach, I., et al. (1990). "Cloning of human hepatic nuclear factor 1 (HNF1) and chromosomal localization of its gene in man and mouse." Genomics **8**(1): 155-164.
- Bacon, S., et al. (2015). "MicroRNA-224 is Readily Detectable in Urine of Individuals with Diabetes Mellitus and is a Potential Indicator of Beta-Cell Demise." Genes (Basel) **6**(2): 399-416.
- Bacon, S., et al. (2016). "Successful maintenance on sulphonylurea therapy and low diabetes complication rates in a HNF1A-MODY cohort." Diabet Med **33**(7): 976-984.
- Bacon, S., et al. (2013). "Circulating CD36 is reduced in HNF1A-MODY carriers." PLoS One **8**(9): e74577.
- Balamurugan, K., et al. (2016). "Structure-function studies of HNF1A (MODY3) gene mutations in South Indian patients with monogenic diabetes." Clin Genet **90**(6): 486-495.
- Bellanne-Chantelot, C., et al. (2008). "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 **57**(2): 503-508.
- Bellanne-Chantelot, C., et al. (2004). "Clinical spectrum associated with hepatocyte nuclear factor-1beta mutations." Ann Intern Med **140**(7): 510-517.

- Bellanne-Chantelot, C., et al. (2005). "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 **54**(11): 3126-3132.
- Bellanne-Chantelot, C., et al. (2016). "High-sensitivity C-reactive protein does not improve the differential diagnosis of HNF1A-MODY and familial young-onset type 2 diabetes: A grey zone analysis." Diabetes Metab **42**(1): 33-37.
- Bellanne-Chantelot, C., et al. (2011). "Clinical characteristics and diagnostic criteria of maturity-onset diabetes of the young (MODY) due to molecular anomalies of the HNF1A gene." J Clin Endocrinol Metab **96**(8): E1346-1351.
- Besser, R. E., et al. (2011). "Urinary C-peptide creatinine ratio is a practical outpatient tool for identifying hepatocyte nuclear factor 1-{alpha}/hepatocyte nuclear factor 4-{alpha} maturity-onset diabetes of the young from long-duration type 1 diabetes." Diabetes Care **34**(2): 286-291.
- Biering, H., et al. (2000). "[Prevalence of diabetes in acromegaly and Cushing syndrome]." Acta Med Austriaca **27**(1): 27-31.
- Bingham, C., et al. (2001). "The generalized aminoaciduria seen in patients with hepatocyte nuclear factor-1alpha mutations is a feature of all patients with diabetes and is associated with glucosuria." Diabetes **50**(9): 2047-2052.
- Bingham, C., et al. (2003). "Atypical familial juvenile hyperuricemic nephropathy associated with a hepatocyte nuclear factor-1beta gene mutation." Kidney Int **63**(5): 1645-1651.
- Biomarkers Definitions Working, G. (2001). "Biomarkers and surrogate endpoints: preferred definitions and conceptual framework." Clin Pharmacol Ther **69**(3): 89-95.
- Bjorkhaug, L., et al. (2003). "Hepatocyte nuclear factor-1 alpha gene mutations and diabetes in Norway." J Clin Endocrinol Metab **88**(2): 920-931.
- Bjorkhaug, L., et al. (2000). "MODY associated with two novel hepatocyte nuclear factor-1alpha loss-of-function mutations (P112L and Q466X)." Biochem Biophys Res Commun **279**(3): 792-798.
- Bluteau, O., et al. (2002). "Bi-allelic inactivation of TCF1 in hepatic adenomas." Nat Genet **32**(2): 312-315.
- Boj, S. F., et al. (2009). "Functional targets of the monogenic diabetes transcription factors HNF-1alpha and HNF-4alpha are highly conserved between mice and humans." Diabetes **58**(5): 1245-1253.
- Bonifacio, E., et al. (2010). "Harmonization of glutamic acid decarboxylase and islet antigen-2 autoantibody assays for national institute of diabetes and digestive and kidney diseases consortia." J Clin Endocrinol Metab **95**(7): 3360-3367.

- Bonnefond, A., et al. (2012). "Whole-exome sequencing and high throughput genotyping identified KCNJ11 as the thirteenth MODY gene." PLoS One **7**(6): e37423.
- Bonner, C., et al. (2013). "Identification of circulating microRNAs in HNF1A-MODY carriers." Diabetologia **56**(8): 1743-1751.
- Boutin, P., et al. (1997). "An automated fluorescent single-strand conformation polymorphism technique for screening mutations in the hepatocyte nuclear factor-1alpha gene (maturity-onset diabetes of the young)." Diabetes **46**(12): 2108-2109.
- Buderer, N. M. (1996). "Statistical methodology: I. Incorporating the prevalence of disease into the sample size calculation for sensitivity and specificity." Acad Emerg Med **3**(9): 895-900.
- Bujalska, I. J., et al. (1997). "Does central obesity reflect "Cushing's disease of the omentum"?" Lancet **349**(9060): 1210-1213.
- Caulfield, M. P., et al. (2002). "The diagnosis of congenital adrenal hyperplasia in the newborn by gas chromatography/mass spectrometry analysis of random urine specimens." J Clin Endocrinol Metab **87**(8): 3682-3690.
- Cervin, C., et al. (2010). "An investigation of serum concentration of apoM as a potential MODY3 marker using a novel ELISA." J Intern Med **267**(3): 316-321.
- Chen, M. and T. M. Penning (2014). "5beta-Reduced steroids and human Delta(4)-3-ketosteroid 5beta-reductase (AKR1D1)." Steroids **83**: 17-26.
- Chen, Y. Z., et al. (2010). "Systematic review of TCF2 anomalies in renal cysts and diabetes syndrome/maturity onset diabetes of the young type 5." Chin Med J (Engl) **123**(22): 3326-3333.
- Chi, Y. I., et al. (2002). "Diabetes mutations delineate an atypical POU domain in HNF-1alpha." Mol Cell **10**(5): 1129-1137.
- Choi, Y., et al. (2012). "Predicting the functional effect of amino acid substitutions and indels." PLoS One **7**(10): e46688.
- Colclough, K., et al. (2013). "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 **34**(5): 669-685.
- Courtois, G., et al. (1987). "Interaction of a liver-specific nuclear factor with the fibrinogen and alpha 1-antitrypsin promoters." Science **238**(4827): 688-692.
- Desmet, F. O., et al. (2009). "Human Splicing Finder: an online bioinformatics tool to predict splicing signals." Nucleic Acids Res **37**(9): e67.
- Domínguez-López A, M.-G. A., Segura-Kato YX, Riba L, Esparza-López R, Ramírez-Jiménez S, Rodríguez-Torres M, Canizales-Quinteros S, Cabrera-Vásquez S, Fragoso-Ontiveros V, Aguilar-Salinas CA, Altamirano-Bustamante N, Calzada-León R, Robles-Valdés C, Bravo-

- Ríos LE, Tusié-Luna MT. (2005). "Mutations in MODY genes are not common cause of early-onset type 2 diabetes in Mexican families." JOP **6**(3): 238-245.
- Dowman, J. K., et al. (2013). "Loss of 5alpha-reductase type 1 accelerates the development of hepatic steatosis but protects against hepatocellular carcinoma in male mice." Endocrinology **154**(12): 4536-4547.
- Edghill, E. L., et al. (2008). "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 **57**(4): 1034-1042.
- Egan, A. M., et al. (2015). "Diabetic ketoacidosis in the setting of HNF1A-maturity onset diabetes of the young." BMJ Case Rep **2015**.
- Eide, S. A., et al. (2008). "Prevalence of HNF1A (MODY3) mutations in a Norwegian population (the HUNT2 Study)." Diabet Med **25**(7): 775-781.
- Ekholm, E., et al. (2013). "Alterations in bile acid synthesis in carriers of hepatocyte nuclear factor 1alpha mutations." J Intern Med **274**(3): 263-272.
- Ellard, S., et al. (2008). "Best practice guidelines for the molecular genetic diagnosis of maturity-onset diabetes of the young." Diabetologia **51**(4): 546-553.
- Ellard, S., et al. (2013). "Improved genetic testing for monogenic diabetes using targeted next-generation sequencing." Diabetologia **56**(9): 1958-1963.
- Ellard, S., et al. (2007). "Partial and whole gene deletion mutations of the GCK and HNF1A genes in maturity-onset diabetes of the young." Diabetologia **50**(11): 2313-2317.
- Estalella, I., et al. (2007). "Mutations in GCK and HNF-1alpha explain the majority of cases with clinical diagnosis of MODY in Spain." Clin Endocrinol (Oxf) **67**(4): 538-546.
- Estrada, K., et al. (2014). "Association of a low-frequency variant in HNF1A with type 2 diabetes in a Latino population." JAMA **311**(22): 2305-2314.
- Fagerberg, L., et al. (2014). "Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics." Mol Cell Proteomics **13**(2): 397-406.
- Fajans, S. S. and G. I. Bell (2011). "MODY: history, genetics, pathophysiology, and clinical decision making." Diabetes Care **34**(8): 1878-1884.
- Fardella, C. E. and W. L. Miller (1996). "Molecular biology of mineralocorticoid metabolism." Annu Rev Nutr **16**: 443-470.
- Feig, P. U., et al. (2011). "Effects of an 11beta-hydroxysteroid dehydrogenase type 1 inhibitor, MK-0916, in patients with type 2 diabetes mellitus and metabolic syndrome." Diabetes Obes Metab **13**(6): 498-504.
- Feigerlova, E., et al. (2006). "Aetiological heterogeneity of asymptomatic hyperglycaemia in children and adolescents." Eur J Pediatr **165**(7): 446-452.

- Finken, M. J., et al. (1999). "Cortisol metabolism in healthy young adults: sexual dimorphism in activities of A-ring reductases, but not 11beta-hydroxysteroid dehydrogenases." J Clin Endocrinol Metab **84**(9): 3316-3321.
- Flannick, J., et al. (2013). "Assessing the phenotypic effects in the general population of rare variants in genes for a dominant Mendelian form of diabetes." Nat Genet **45**(11): 1380-1385.
- Frayling, T. M., et al. (2001). "beta-cell genes and diabetes - Molecular and clinical characterization of mutations in transcription factors." Diabetes **50**: S94-S100.
- Freeze, H. H. (2006). "Genetic defects in the human glycome." Nat Rev Genet **7**(7): 537-551.
- Friedman, T. C., et al. (1996). "Carbohydrate and lipid metabolism in endogenous hypercortisolism: shared features with metabolic syndrome X and NIDDM." Endocr J **43**(6): 645-655.
- Fuchsberger, C., et al. (2016). "The genetic architecture of type 2 diabetes." Nature **536**(7614): 41-47.
- Galan, M., et al. (2011). "Differential Effects of HNF-1 alpha Mutations Associated with Familial Young-Onset Diabetes on Target Gene Regulation." Molecular Medicine **17**(3-4): 256-265.
- Gathercole, L. L., et al. (2011). "Regulation of lipogenesis by glucocorticoids and insulin in human adipose tissue." PLoS One **6**(10): e26223.
- Gilliam, L. K., et al. (2013). "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.
- Gloyn, A. L., et al. (2012). "Metabolic profiling in Maturity-onset diabetes of the young (MODY) and young onset type 2 diabetes fails to detect robust urinary biomarkers." PLoS One **7**(7): e40962.
- Glucksmann, M. A., et al. (1997). "Novel mutations and a mutational hotspot in the MODY3 gene." Diabetes **46**(6): 1081-1086.
- Goring, H. H., et al. (2001). "Large upward bias in estimation of locus-specific effects from genomewide scans." Am J Hum Genet **69**(6): 1357-1369.
- Gornik, O., et al. (2012). "Alternative glycosylation modulates function of IgG and other proteins - implications on evolution and disease." Biochim Biophys Acta **1820**(9): 1318-1326.
- Gosetti, F., et al. (2013). "Ultra high performance liquid chromatography tandem mass spectrometry determination and profiling of prohibited steroids in human biological matrices. A review." J Chromatogr B Analyt Technol Biomed Life Sci **927**: 22-36.

Goto, Y., et al. (1990). "A mutation in the tRNA(Leu)(UUR) gene associated with the MELAS subgroup of mitochondrial encephalomyopathies." Nature **348**(6302): 651-653.

Guazzarotti, L., et al. (1999). "Maturity-onset diabetes of the young (MODY): a new challenge for pediatric diabetologists." J Pediatr Endocrinol Metab **12**(4): 487-497.

Guillausseau, P. J., et al. (2001). "Maternally inherited diabetes and deafness: a multicenter study." Ann Intern Med **134**(9 Pt 1): 721-728.

Gupta, R. K., et al. (2005). "The MODY1 gene HNF-4alpha regulates selected genes involved in insulin secretion." J Clin Invest **115**(4): 1006-1015.

Hansen, S. K., et al. (2002). "Genetic evidence that HNF-1alpha-dependent transcriptional control of HNF-4alpha is essential for human pancreatic beta cell function." J Clin Invest **110**(6): 827-833.

Harries, L. W., et al. (2006). "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 **15**(14): 2216-2224.

Harries, L. W., et al. (2008). "Diabetes susceptibility in the Canadian Oji-Cree population is moderated by abnormal mRNA processing of HNF1A G319S transcripts." Diabetes **57**(7): 1978-1982.

Hartemann-Heurtier, A., et al. (2009). "Mutations in the ABCC8 gene can cause autoantibody-negative insulin-dependent diabetes." Diabetes Metab **35**(3): 233-235.

Hassan-Smith, Z. K., et al. (2015). "Gender-Specific Differences in Skeletal Muscle 11beta-HSD1 Expression Across Healthy Aging." J Clin Endocrinol Metab **100**(7): 2673-2681.

Hattersley, A. T., et al. (1992). "Linkage of type 2 diabetes to the glucokinase gene." Lancet **339**(8805): 1307-1310.

Hazlehurst, J. M., et al. (2016). "Dual-5alpha-Reductase Inhibition Promotes Hepatic Lipid Accumulation in Man." J Clin Endocrinol Metab **101**(1): 103-113.

Hegele, R. A., et al. (1999). "The hepatic nuclear factor-1alpha G319S variant is associated with early-onset type 2 diabetes in Canadian Oji-Cree." J Clin Endocrinol Metab **84**(3): 1077-1082.

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

Huffman, J. E., et al. (2014). "Comparative Performance of Four Methods for High-throughput Glycosylation Analysis of Immunoglobulin G in Genetic and Epidemiological Research." Molecular & Cellular Proteomics **13**(6): 1598-1610.

Iafusco, D., et al. (2002). "Permanent diabetes mellitus in the first year of life." Diabetologia **45**(6): 798-804.

- Isomaa, B., et al. (1998). "Chronic diabetic complications in patients with MODY3 diabetes." Diabetologia **41**(4): 467-473.
- Iwanishi, M., et al. (1995). "Clinical and laboratory characteristics in the families with diabetes and a mitochondrial tRNA(LEU(UUR)) gene mutation." Diabetes Res Clin Pract **29**(2): 75-82.
- Javaud, C., et al. (2003). "The fucosyltransferase gene family: an amazing summary of the underlying mechanisms of gene evolution." Genetica **118**(2-3): 157-170.
- Jitrapakdee, S. (2012). "Transcription factors and coactivators controlling nutrient and hormonal regulation of hepatic gluconeogenesis." Int J Biochem Cell Biol **44**(1): 33-45.
- Juszczak A, T. G., Khonsari M, Webster AL, Owen KR (2014). Diagnostic and therapeutic challenge in a patient with mixed phenotype diabetes. Diabetes UK, Liverpool, UK, Diabetic Medicine
- Juszczak, A. O., K. (2014). "Identifying subtypes of monogenic diabetes." Diabetes Management **4** (1): 713-722.
- Kalia, S. S., et al. (2017). "Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2.0): a policy statement of the American College of Medical Genetics and Genomics." Genet Med **19**(2): 249-255.
- Karlsson, E., et al. (2008). "Can complement factors 5 and 8 and transthyretin be used as biomarkers for MODY 1 (HNF4A-MODY) and MODY 3 (HNF1A-MODY)?" Diabet Med **25**(7): 788-791.
- Kavvoura, F. K., et al. (2014). "Reclassification of diabetes etiology in a family with multiple diabetes phenotypes." J Clin Endocrinol Metab **99**(6): E1067-1071.
- Kawamoto, T., et al. (1992). "Role of steroid 11 beta-hydroxylase and steroid 18-hydroxylase in the biosynthesis of glucocorticoids and mineralocorticoids in humans." Proc Natl Acad Sci U S A **89**(4): 1458-1462.
- Knebel, B., et al. (2016). "Divergent phenotypes in siblings with identical novel mutations in the HNF-1alpha gene leading to maturity onset diabetes of the young type 3." BMC Med Genet **17**(1): 36.
- Krone, N., et al. (2010). "Gas chromatography/mass spectrometry (GC/MS) remains a pre-eminent discovery tool in clinical steroid investigations even in the era of fast liquid chromatography tandem mass spectrometry (LC/MS/MS)." J Steroid Biochem Mol Biol **121**(3-5): 496-504.
- Kropff, J., et al. (2011). "Prevalence of monogenic diabetes in young adults: a community-based, cross-sectional study in Oxfordshire, UK." Diabetologia **54**(5): 1261-1263.

Kumar, P., et al. (2009). "Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm." Nat Protoc **4**(7): 1073-1081.

Lambert, A. P., et al. (2003). "Identifying hepatic nuclear factor 1alpha mutations in children and young adults with a clinical diagnosis of type 1 diabetes." Diabetes Care **26**(2): 333-337.

Lauc, G., et al. (2010). "Genomics meets glycomics-the first GWAS study of human N-Glycome identifies HNF1alpha as a master regulator of plasma protein fucosylation." PLoS Genet **6**(12): e1001256.

Ledermann, H. M. (1995). "Maturity-onset diabetes of the young (MODY) at least ten times more common in Europe than previously assumed?" Diabetologia **38**(12): 1482.

Lehto, M., et al. (1997). "Characterization of the MODY3 phenotype - Early-onset diabetes caused by an insulin secretion defect." Journal of Clinical Investigation **99**(4): 582-591.

Lek, M., et al. (2016). "Analysis of protein-coding genetic variation in 60,706 humans." Nature **536**(7616): 285-291.

Lorini, R., et al. (2009). "Maturity-onset diabetes of the young in children with incidental hyperglycemia: a multicenter Italian study of 172 families." Diabetes Care **32**(10): 1864-1866.

MacArthur, D. G., et al. (2012). "A systematic survey of loss-of-function variants in human protein-coding genes." Science **335**(6070): 823-828.

MacArthur, D. G., et al. (2014). "Guidelines for investigating causality of sequence variants in human disease." Nature **508**(7497): 469-476.

Malecki, M. T., et al. (1999). "Mutations in NEUROD1 are associated with the development of type 2 diabetes mellitus." Nat Genet **23**(3): 323-328.

Martin, D., et al. (2008). "Long-term follow-up of oral glucose tolerance test-derived glucose tolerance and insulin secretion and insulin sensitivity indexes in subjects with glucokinase mutations (MODY2)." Diabetes Care **31**(7): 1321-1323.

McCarthy, M. I. (2017). "Painting a new picture of personalised medicine for diabetes." Diabetologia **60**(5): 793-799.

McDonald, T. J., et al. (2011). "Islet autoantibodies can discriminate maturity-onset diabetes of the young (MODY) from Type 1 diabetes." Diabet Med **28**(9): 1028-1033.

McDonald, T. J., et al. (2012). "Lipoprotein composition in HNF1A-MODY: differentiating between HNF1A-MODY and type 2 diabetes." Clin Chim Acta **413**(9-10): 927-932.

Mendel, D. B. and G. R. Crabtree (1991). "HNF-1, a member of a novel class of dimerizing homeodomain proteins." J Biol Chem **266**(2): 677-680.

Miller, W. L. (1988). "Molecular biology of steroid hormone synthesis." Endocr Rev **9**(3): 295-318.

- Miller, W. L. and R. J. Auchus (2011). "The molecular biology, biochemistry, and physiology of human steroidogenesis and its disorders." Endocr Rev **32**(1): 81-151.
- Molven, A. and P. R. Njolstad (2011). "Role of molecular genetics in transforming diagnosis of diabetes mellitus." Expert Rev Mol Diagn **11**(3): 313-320.
- Morris, A. P., et al. (2012). "Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes." Nat Genet **44**(9): 981-990.
- Mrozinska, S., et al. (2016). "Qualitative Parameters of the Colonic Flora in Patients with HNF1A-MODY Are Different from Those Observed in Type 2 Diabetes Mellitus." J Diabetes Res **2016**: 3876764.
- Mughal, S. A., et al. (2013). "Apolipoprotein M can discriminate HNF1A-MODY from Type 1 diabetes." Diabet Med **30**(2): 246-250.
- Murphy, R., et al. (2008). "Clinical features, diagnosis and management of maternally inherited diabetes and deafness (MIDD) associated with the 3243A>G mitochondrial point mutation." Diabet Med **25**(4): 383-399.
- Najmi, L. A., et al. (2016). "Functional Investigations of HNF1A Identify Rare Variants as Risk Factors for Type 2 Diabetes in the General Population." Diabetes.
- Nanus, D. E., et al. (2015). "Differential glucocorticoid metabolism in patients with persistent versus resolving inflammatory arthritis." Arthritis Res Ther **17**: 121.
- Narayana, N., et al. (2001). "The dimerization domain of HNF-1 α : structure and plasticity of an intertwined four-helix bundle with application to diabetes mellitus." J Mol Biol **310**(3): 635-658.
- Naylor, R. N., et al. (2014). "Cost-effectiveness of MODY genetic testing: translating genomic advances into practical health applications." Diabetes Care **37**(1): 202-209.
- Ng, M. C., et al. (1999). "Molecular genetics of diabetes mellitus in Chinese subjects: identification of mutations in glucokinase and hepatocyte nuclear factor-1 α genes in patients with early-onset type 2 diabetes mellitus/MODY." Diabet Med **16**(11): 956-963.
- Nicosia, A., et al. (1990). "A myosin-like dimerization helix and an extra-large homeodomain are essential elements of the tripartite DNA binding structure of LFB1." Cell **61**(7): 1225-1236.
- Nowak, N., et al. (2013). "Cystatin C is not a good candidate biomarker for HNF1A-MODY." Acta Diabetol **50**(5): 815-820.
- Odom, D. T., et al. (2004). "Control of pancreas and liver gene expression by HNF transcription factors." Science **303**(5662): 1378-1381.
- Okada, Y., et al. (2011). "Genome-wide association study for C-reactive protein levels identified pleiotropic associations in the IL6 locus." Hum Mol Genet **20**(6): 1224-1231.

- Olefsky, J. M., et al. (1975). "The effects of acute and chronic dexamethasone administration on insulin binding to isolated rat hepatocytes and adipocytes." Metabolism **24**(4): 517-527.
- Osbak, K. K., et al. (2009). "Update on mutations in glucokinase (GCK), which cause maturity-onset diabetes of the young, permanent neonatal diabetes, and hyperinsulinemic hypoglycemia." Hum Mutat **30**(11): 1512-1526.
- Ovsyannikova, A. K., et al. (2016). "ABCC8-Related Maturity-Onset Diabetes of the Young (MODY12): Clinical Features and Treatment Perspective." Diabetes Ther **7**(3): 591-600.
- Owen, K. R., et al. (2002). "Heterogeneity in young adult onset diabetes: aetiology alters clinical characteristics." Diabet Med **19**(9): 758-761.
- Owen, K. R., et al. (2010). "Assessment of high-sensitivity C-reactive protein levels as diagnostic discriminator of maturity-onset diabetes of the young due to HNF1A mutations." Diabetes Care **33**(9): 1919-1924.
- Palermo, M., et al. (2004). "Apparent mineralocorticoid excess syndrome: an overview." Arq Bras Endocrinol Metabol **48**(5): 687-696.
- Palermo, M., et al. (1996). "Urinary free cortisone and the assessment of 11 beta-hydroxysteroid dehydrogenase activity in man." Clin Endocrinol (Oxf) **45**(5): 605-611.
- Panzram, G. and W. Adolph (1981). "Heterogeneity of maturity onset diabetes at young age (MODY)." Lancet **2**(8253): 986.
- Pearson, E. R., et al. (2004). "Contrasting diabetes phenotypes associated with hepatocyte nuclear factor-1alpha and -1beta mutations." Diabetes Care **27**(5): 1102-1107.
- Pearson, E. R., et al. (2007). "Macrosomia and hyperinsulinaemic hypoglycaemia in patients with heterozygous mutations in the HNF4A gene." PLoS Med **4**(4): e118.
- Pearson, E. R., et al. (2005). "Molecular genetics and phenotypic characteristics of MODY caused by hepatocyte nuclear factor 4alpha mutations in a large European collection." Diabetologia **48**(5): 878-885.
- Pearson, E. R., et al. (2003). "Genetic cause of hyperglycaemia and response to treatment in diabetes." Lancet **362**(9392): 1275-1281.
- Plengvidhya, N., et al. (2007). "PAX4 mutations in Thais with maturity onset diabetes of the young." J Clin Endocrinol Metab **92**(7): 2821-2826.
- Pontoglio, M., et al. (1996). "Hepatocyte nuclear factor 1 inactivation results in hepatic dysfunction, phenylketonuria, and renal Fanconi syndrome." Cell **84**(4): 575-585.
- Pontoglio, M., et al. (2000). "HNF1alpha controls renal glucose reabsorption in mouse and man." EMBO Rep **1**(4): 359-365.
- Pontoglio, M., et al. (1998). "Defective insulin secretion in hepatocyte nuclear factor 1alpha-deficient mice." J Clin Invest **101**(10): 2215-2222.

Qin, K. N., et al. (1993). "Molecular cloning of multiple cDNAs encoding human enzymes structurally related to 3 alpha-hydroxysteroid dehydrogenase." J Steroid Biochem Mol Biol **46**(6): 673-679.

Quintans, B., et al. (2014). "Medical genomics: The intricate path from genetic variant identification to clinical interpretation." Appl Transl Genom **3**(3): 60-67.

Rask, E., et al. (2002). "Tissue-specific changes in peripheral cortisol metabolism in obese women: increased adipose 11beta-hydroxysteroid dehydrogenase type 1 activity." J Clin Endocrinol Metab **87**(7): 3330-3336.

Reihner, E., et al. (1989). "Bile acid synthesis in humans: regulation of hepatic microsomal cholesterol 7 alpha-hydroxylase activity." Gastroenterology **97**(6): 1498-1505.

Reiner, A. P., et al. (2008). "Polymorphisms of the HNF1A gene encoding hepatocyte nuclear factor-1 alpha are associated with C-reactive protein." Am J Hum Genet **82**(5): 1193-1201.

Reznik, Y., et al. (2004). "Hepatocyte nuclear factor-1 alpha gene inactivation: cosegregation between liver adenomatosis and diabetes phenotypes in two maturity-onset diabetes of the young (MODY)3 families." J Clin Endocrinol Metab **89**(3): 1476-1480.

Ricketts, M. L., et al. (1998). "Immunohistochemical localization of type 1 11beta-hydroxysteroid dehydrogenase in human tissues." J Clin Endocrinol Metab **83**(4): 1325-1335.

Ridker, P. M., et al. (2008). "Loci related to metabolic-syndrome pathways including LEPR, HNF1A, IL6R, and GCKR associate with plasma C-reactive protein: the Women's Genome Health Study." Am J Hum Genet **82**(5): 1185-1192.

Rivas, M. A., et al. (2015). "Human genomics. Effect of predicted protein-truncating genetic variants on the human transcriptome." Science **348**(6235): 666-669.

Rooks, J. B., et al. (1979). "Epidemiology of hepatocellular adenoma. The role of oral contraceptive use." JAMA **242**(7): 644-648.

Rose, A. S. and P. W. Hildebrand (2015). "NGL Viewer: a web application for molecular visualization." Nucleic Acids Res **43**(W1): W576-579.

Rose, R. B., et al. (2000). "Structural basis of dimerization, coactivator recognition and MODY3 mutations in HNF-1alpha." Nat Struct Biol **7**(9): 744-748.

Rosenstock, J., et al. (2010). "The 11-beta-hydroxysteroid dehydrogenase type 1 inhibitor INCB13739 improves hyperglycemia in patients with type 2 diabetes inadequately controlled by metformin monotherapy." Diabetes Care **33**(7): 1516-1522.

Russell, D. W. and J. D. Wilson (1994). "Steroid 5 alpha-reductase: two genes/two enzymes." Annu Rev Biochem **63**: 25-61.

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

Saad, M. J., et al. (1993). "Modulation of insulin receptor, insulin receptor substrate-1, and phosphatidylinositol 3-kinase in liver and muscle of dexamethasone-treated rats." J Clin Invest **92**(4): 2065-2072.

Salдова, R., et al. (2014). "Association of N-glycosylation with breast carcinoma and systemic features using high-resolution quantitative UPLC." J Proteome Res **13**(5): 2314-2327.

Sanger, F., et al. (1977). "DNA sequencing with chain-terminating inhibitors." Proc Natl Acad Sci U S A **74**(12): 5463-5467.

Scherer, W. F., et al. (1953). "Studies on the propagation in vitro of poliomyelitis viruses. IV. Viral multiplication in a stable strain of human malignant epithelial cells (strain HeLa) derived from an epidermoid carcinoma of the cervix." J Exp Med **97**(5): 695-710.

Schober, E., et al. (2009). "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 **26**(5): 466-473.

Schwarz, J. M., et al. (2014). "MutationTaster2: mutation prediction for the deep-sequencing age." Nat Methods **11**(4): 361-362.

Scott, R. A., et al. (2017). "An Expanded Genome-Wide Association Study of Type 2 Diabetes in Europeans." Diabetes.

Servitja, J. M., et al. (2009). "Hnf1alpha (MODY3) controls tissue-specific transcriptional programs and exerts opposed effects on cell growth in pancreatic islets and liver." Mol Cell Biol **29**(11): 2945-2959.

Shackleton, C. H. (1986). "Profiling steroid hormones and urinary steroids." J Chromatogr **379**: 91-156.

Shah, N., et al. (2014). "Comparability of high-sensitivity CRP methods to detect maturity-onset diabetes of the young due to HNF1A mutations." Br J Biomed Sci **71**(2): 84-85.

Shalem, O., et al. (2015). "High-throughput functional genomics using CRISPR-Cas9." Nat Rev Genet **16**(5): 299-311.

Shepherd M, S. A., and Hattersley AT (2001). "Genetic testing in maturity onset diabetes of the young (MODY): a new challenge for the diabetic clinic." Pract Diab Int **18**: 16-21.

Shepherd, M., et al. (2009). "A genetic diagnosis of HNF1A diabetes alters treatment and improves glycaemic control in the majority of insulin-treated patients." Diabet Med **26**(4): 437-441.

Shields, B. M., et al. (2010). "Maturity-onset diabetes of the young (MODY): how many cases are we missing?" Diabetologia **53**(12): 2504-2508.

Shields, B. M., et al. (2012). "The development and validation of a clinical prediction model to determine the probability of MODY in patients with young-onset diabetes." Diabetologia **55**(5): 1265-1272.

Skupien, J., et al. (2008). "Clinical application of 1,5-anhydroglucitol measurements in patients with hepatocyte nuclear factor-1alpha maturity-onset diabetes of the young." Diabetes Care **31**(8): 1496-1501.

Socas, L., et al. (2005). "Hepatocellular adenomas associated with anabolic androgenic steroid abuse in bodybuilders: a report of two cases and a review of the literature." Br J Sports Med **39**(5): e27.

Spyer, G., et al. (2001). "Influence of maternal and fetal glucokinase mutations in gestational diabetes." Am J Obstet Gynecol **185**(1): 240-241.

Starita, L. M., et al. (2017). "Variant Interpretation: Functional Assays to the Rescue." Am J Hum Genet **101**(3): 315-325.

Steele, A. M., et al. (2010). "Increased all-cause and cardiovascular mortality in monogenic diabetes as a result of mutations in the HNF1A gene." Diabet Med **27**(2): 157-161.

Steele, A. M., et al. (2014). "Prevalence of vascular complications among patients with glucokinase mutations and prolonged, mild hyperglycemia." JAMA **311**(3): 279-286.

Steele, A. M., et al. (2013). "Use of HbA1c in the Identification of Patients with Hyperglycaemia Caused by a Glucokinase Mutation: Observational Case Control Studies." PLoS One **8**(6): e65326.

Stoffers, D. A., et al. (1997). "Early-onset type-II diabetes mellitus (MODY4) linked to IPF1." Nat Genet **17**(2): 138-139.

Stride, A., et al. (2005). "Beta-cell dysfunction, insulin sensitivity, and glycosuria precede diabetes in hepatocyte nuclear factor-1alpha mutation carriers." Diabetes Care **28**(7): 1751-1756.

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

Stride, A., et al. (2002). "The genetic abnormality in the beta cell determines the response to an oral glucose load." Diabetologia **45**(3): 427-435.

Tattersall, R. B. (1974). "Mild familial diabetes with dominant inheritance." Q J Med **43**(170): 339-357.

Teo, A. K., et al. (2013). "Derivation of human induced pluripotent stem cells from patients with maturity onset diabetes of the young." J Biol Chem **288**(8): 5353-5356.

Thanabalasingham, G., et al. (2013). "Mutations in HNF1A result in marked alterations of plasma glycan profile." Diabetes **62**(4): 1329-1337.

Thanabalasingham, G., et al. (2012). "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 **35**(6): 1206-1212.

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

Tillil, H. and J. Kobberling (1987). "Age-corrected empirical genetic risk estimates for first-degree relatives of IDDM patients." Diabetes **36**(1): 93-99.

Timar, O., et al. (2000). "Metabolic syndrome X: a review." Can J Cardiol **16**(6): 779-789.

Tomlinson, J. W., et al. (2008). "Impaired glucose tolerance and insulin resistance are associated with increased adipose 11beta-hydroxysteroid dehydrogenase type 1 expression and elevated hepatic 5alpha-reductase activity." Diabetes **57**(10): 2652-2660.

Tomlinson, J. W., et al. (2004). "11beta-hydroxysteroid dehydrogenase type 1: a tissue-specific regulator of glucocorticoid response." Endocr Rev **25**(5): 831-866.

Toniatti, C., et al. (1990). "Synergistic trans-activation of the human C-reactive protein promoter by transcription factor HNF-1 binding at two distinct sites." EMBO J **9**(13): 4467-4475.

Trajkovski, M., et al. (2011). "MicroRNAs 103 and 107 regulate insulin sensitivity." Nature **474**(7353): 649-653.

Triggs-Raine, B. L., et al. (2002). "HNF-1alpha G319S, a transactivation-deficient mutant, is associated with altered dynamics of diabetes onset in an Oji-Cree community." Proc Natl Acad Sci U S A **99**(7): 4614-4619.

Tuomi, T., et al. (2006). "Improved prandial glucose control with lower risk of hypoglycemia with nateglinide than with glibenclamide in patients with maturity-onset diabetes of the young type 3." Diabetes Care **29**(2): 189-194.

Uhlen, M., et al. (2015). "Proteomics. Tissue-based map of the human proteome." Science **347**(6220): 1260419.

Valsamakis, G., et al. (2004). "11beta-hydroxysteroid dehydrogenase type 1 activity in lean and obese males with type 2 diabetes mellitus." J Clin Endocrinol Metab **89**(9): 4755-4761.

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

Vaxillaire, M., et al. (1995). "A gene for maturity onset diabetes of the young (MODY) maps to chromosome 12q." Nat Genet **9**(4): 418-423.

- Velho, G., et al. (1997). "Identification of 14 new glucokinase mutations and description of the clinical profile of 42 MODY-2 families." Diabetologia **40**(2): 217-224.
- Velho, G., et al. (2000). "Maternal diabetes alters birth weight in glucokinase-deficient (MODY2) kindred but has no influence on adult weight, height, insulin secretion or insulin sensitivity." Diabetologia **43**(8): 1060-1063.
- Verge, C. F., et al. (1994). "Anti-glutamate decarboxylase and other antibodies at the onset of childhood IDDM: a population-based study." Diabetologia **37**(11): 1113-1120.
- Voight, B. F., et al. (2010). "Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis." Nat Genet **42**(7): 579-589.
- Wallace, T. M., et al. (2004). "Use and abuse of HOMA modeling." Diabetes Care **27**(6): 1487-1495.
- Walsh, R., et al. (2017). "Reassessment of Mendelian gene pathogenicity using 7,855 cardiomyopathy cases and 60,706 reference samples." Genet Med **19**(2): 192-203.
- Whiting, P., et al. (2004). "Sources of variation and bias in studies of diagnostic accuracy: a systematic review." Ann Intern Med **140**(3): 189-202.
- Wooding, K. M. and R. J. Auchus (2013). "Mass spectrometry theory and application to adrenal diseases." Mol Cell Endocrinol **371**(1-2): 201-207.
- Yamada, S., et al. (1999). "Identification of mutations in the hepatocyte nuclear factor-1alpha gene in Japanese subjects with early-onset NIDDM and functional analysis of the mutant proteins." Diabetes **48**(3): 645-648.
- Yamagata, K., et al. (1996). "Mutations in the hepatocyte nuclear factor-1alpha gene in maturity-onset diabetes of the young (MODY3)." Nature **384**(6608): 455-458.
- Yorifuji, T., et al. (2005). "The C42R mutation in the Kir6.2 (KCNJ11) gene as a cause of transient neonatal diabetes, childhood diabetes, or later-onset, apparently type 2 diabetes mellitus." J Clin Endocrinol Metab **90**(6): 3174-3178.