JOSIP JURAJ STROSSMAYER UNIVERSITY OSIJEK
UNIVERSITY OF DUBROVNIK
THE RUĐER BOŠKOVIĆ INSTITUTE

UNIVERSITY POSTGRADUATE
INTERDISCIPLINARY DOCTORAL STUDY
MOLECULAR BIOSCIENCES

MISLAV NOVOKMET

The role of transcription factor HNF1A in glycosylation of plasma proteins

A DISSERTATION

OSIJEK, 2013

TEMELJNA DOKUMENTACIJSKA KARTICA

Sveučilište Josipa Jurja Strossmayera u Osijeku **Doktorski rad**
Institut Ruđer Bošković, Zagreb
Sveučilište u Dubrovniku
Sveučilišni poslijediplomski interdisciplinarni doktorski studij Molekularne bioznanosti

Znanstveno područje: Prirodne znanosti
Znanstveno polje: Biologija

Uloga transkripcijskog faktora HNF1A u glikozilaciji plazmatskih proteina

Mislav Novokmet

Rad je izrađen u: Genos d.o.o.
Mentor: Dr.sc. Olga Gornik, doc.

Sažetak

Gotovo svi plazmatski proteini modificirani su raznolikim glikanskim strukturama koje imaju ključnu ulogu za većinu funkcija ovih proteina. Promjene sastava glikanskih struktura primijećene su u mnogim patološkim stanjima te se istražuju kao potencijalni dijagnostički biljezi. Pokazana je nova uloga hepatocitnog nuklearnog faktora faktor 1A (HNF1A) kao regulatora ključnih koraka u procesu fukozilacije. Mutacije unutar gena za HNF1A koje su odgovorne za HNF1A-MODY (*maturity.onset diabetes oft the young*), jedan od podtipova dijabetesa, utječu na poremećenu fukozilaciju plazmatskih proteina. Između pacijenata oboljelih od HNF1A-MODY-a i pacijenata s drugim podtipovima dijabetesa postoji značajna razlika u razini antenarne fukozilacije glikanskih struktura što predstavlja novi potencijalni biomarker za dijagnosticiranja HNF1A-MODY-a.

Broj stranica: 92
Broj slika: 31
Broj tablica: 23
Broj literaturnih navoda: 94
Jezik izvornika: Engleski

Ključne riječi: Glikoproteini, glikozilacija, fukozilacija, HNF1A, MODY, biomarker

Datum obrane: 02. svibnja 2013.

Stručno povjerenstvo za obranu:

1.
2.
3.

Rad je pohranjen u:

Nacionalnoj i sveučilišnoj knjižnici u Zagrebu, Hrvatske bratske zajednice bb, Zagreb Sveučilištu u Zagrebu, Trg maršala Tita 14, Zagreb
Sveučilištu u Rijeci, Riječke rezolucije 7, Rijeka
Sveučilištu u Splitu, Livanjska 5, Split
Sveučilištu Josipa Jurja Strossmayera u Osijeku, Trg Sv.Trojstva 3, Osijek

BASIC DOCUMENTATION CARD

University Josip Juraj Strossmayer Osijek **PhD thesis**
Institute Ruđer Bošković, Zagreb
University of Dubrovnik
University Postgraduate Interdisciplinary
Doctoral Study Molecular Biosciences

Scientific Area: Natural sciences
Scientific Field: Biology

The role of transcription factor HNF1A in glycosylation of plasma proteins

Mislav Novokmet

Thesis performed at Genos Ltd.
Supervisor: Olga Gornik, PhD

Abstract

Glycosylation is the most diverse posttranslational protein modification. Over half of all known proteins are modified by covalently bound glycans, which are important for normal protein function. In numerous pathological conditions levels of a particular glycan structures are changing independently of protein concentration. A new role of the hepatocyte nuclear factor (HNF)-1A gene is shown as the transcriptional regulator of plasma protein fucosylation. Mutations in the HNF1A gene responsible for the subtype of diabetes HNF1A-MODY (maturity onset diabetes of the young) are also responsible for the decreased plasma protein antennary fucosylation. Numerous mutations, spread across ten exons, may cause inactivation of HNF1A transcription factor and subsequently lower levels of plasma protein fucosylation which is a promising biomarker for both, a HNF1A-MODY diagnosis and HNF1A impaired function.

Number of pages: 92
Number of figures: 31
Number of tables: 23
Number of references: 94
Original in: English

Key words: Glycoproteins, glycosylation, fucosylation, HNF1A, MODY, biomarker

Date of the thesis defense: 02. May 2013.

Reviewers:

1.

2.

3.

Thesis deposited in:

National and University Library, Hrvatske bratske zajednice bb, Zagreb
University of Zagreb, Trg maršala Tita 14, Zagreb
University of Rijeka, Riječke rezolucije 7, Rijeka
University of Split, Livanjska 5, Split
University Josip Juraj Strossmayer Osijek, Trg Sv. Trojstva 3, Osijek

This study was done at Glycobiology Laboratory, Genos Ltd., Zagreb, under the supervision of Assistant Professor Olga Gornik, PhD. The research was funded by FP7 grant GlycoBiom led by Professor Gordan Lauc, PhD.

Many thanks to my advisor Olga Gornik for the support of my study and everyday research.

I thank Professor Gordan Lauc for the opportunity to do my PhD at Genos.

I would like to thank all our numerous collaborators, especially to team at Oxford Centre for Diabetes, Endocrinology & Metabolism and people at Institute of Genetics and Molecular Medicine, University of Edinburgh.

Sincere thanks to Genos team without whom this work would not be possible, especially glyco-pioneers, Maja, Ana and Irma.

Thank you FRano

Many thanks to my friends, nothing seems hard after few moments with them.

And finally, I thank my family who supported me throughout my life.

Table of contents

1 INTRODUCTION

1.1 GLYCOSYLATION

The majority of all membrane and secreted proteins[1], as well as numerous cytoplasmic proteins[2,3], have one or several specific branched oligosaccharide chains (glycans) attached to their backbone. Those proteins are referred to as glycoproteins and the process of oligosaccharide attachment to a protein is called glycosylation. Glycans have numerous important structural, functional and regulatory roles in various physiological processes including protein degradation, folding and secretion, cell signaling, immune function and transcription[4,5,6]. Glycans on the cell surface are the primary attachment site for the majority microorganisms[7] and in fact all interactions that take place at the cell surface are modulated by glycans[8].

While structure of the polypeptide part of a glycoprotein is defined by the sequence of nucleotides in the corresponding gene, structure of a glycan part results from dynamic interactions between hundreds of genes, their protein products and environmental factors[6]. The composition of the glycome (all glycan structures that encompass an organism) attached to an individual protein, or to a complex mixture of proteins, like human plasma, is stable within an individual, but very variable between individuals[9]. This variability stems from numerous common genetic polymorphisms reflecting in changes in the complex biosynthetic pathway of glycans, but also from the interaction with the environment[10]. Mutations that interfere with the basic setup of protein glycosylation machinery are deleterious and are frequently embryonically lethal[11]. However, mutations in genes involved in modifications of glycan antennas are common and apparently cause a large part of individual phenotypic variations that exist in humans and other higher organisms. The most prominent example is the AB0 system of blood group determination, which is based on the existence of three allelic variants of a single glycosyltransferase gene[12]. The majority of human variability originates from single nucleotide polymorphisms (SNPs) that individually do not have visible phenotypes, but if present in specific combinations, they can have significant phenotypic effects[13].

Mammals have evolved a complex glycan repertoire which is estimated to be around thousands of glycan structures and could be larger than the proteome[14].

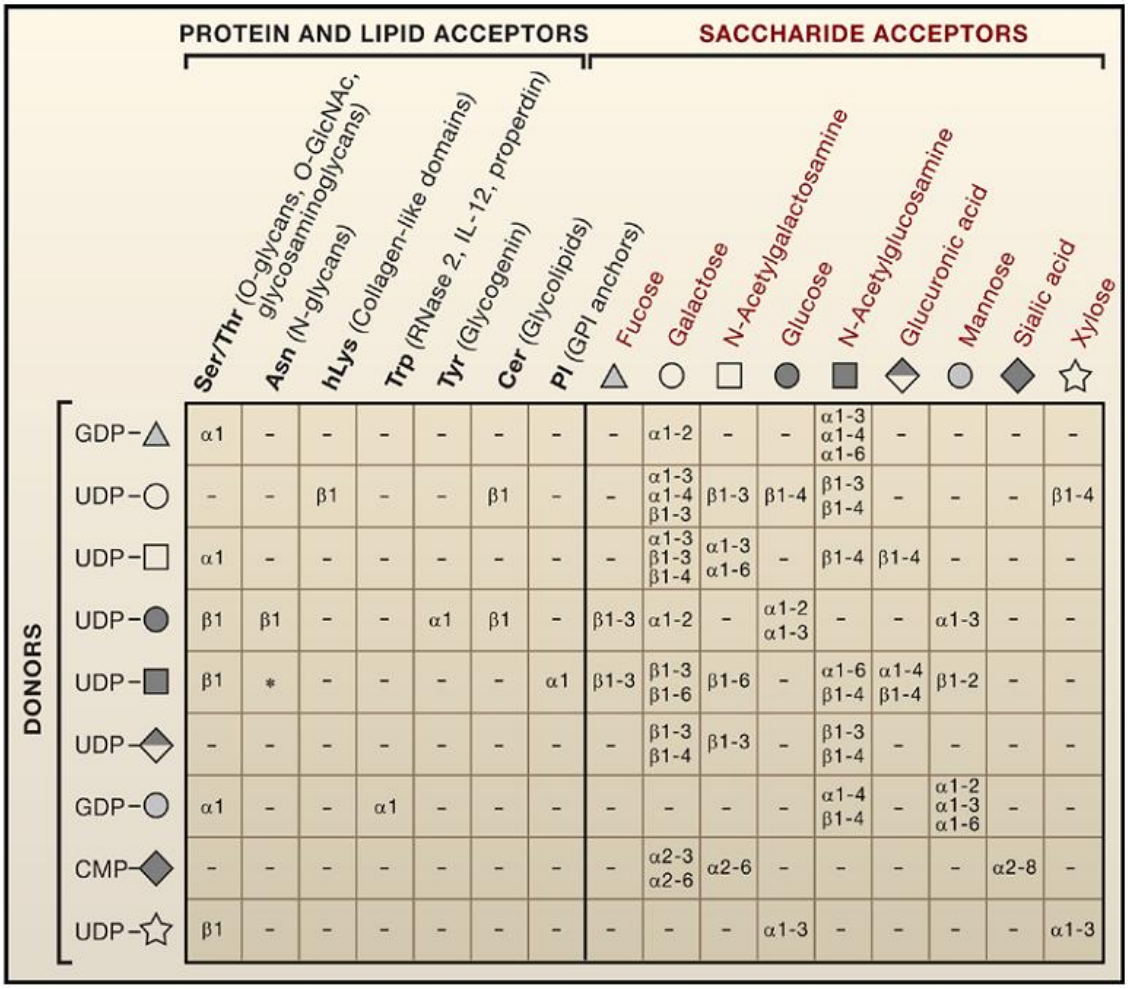


Figure 1: Nine nucleotide sugar donors, protein and lipid acceptor motifs and glycan linkages produced by the complex biosynthetic pathway of mammals[14].

This complex mammalian glycome is built from nine monosaccharides which makes 10^{12} theoretical combinations of hexasaccharides possible (Figure 1). Constrains are provided by the glycosylation machinery, a portfolio of cellular enzymes and substrates which allow only a portion of theoretical combinations. Differences in monosaccharide composition, anomeric state, linkage of the subunits, branching and linkage to the peptide part of a glycoprotein are all contributing to the diversity of the glycan portion of the glycoprotein [14,15]. Expression of human glycome is regulated at both a posttranscriptional and posttranslational level[14]. It involves the control of intracellular localization, catalytic regulation, substrate competition and trafficking control between the endoplasmatic reticulum and Golgi [16,17]. According to protein linkage, glycans can be divided in N- and O-glycans. O-glycans are covalently

attached to serine or threonine residues and N-glycans at aspargine residues of the protein by an N- glycosidic bond. Only the N-glycosylation analysis of the human plasma proteins was carried out in the scope of this work.

1.2 N – GLYCOSYLATION

N-glycans share a common core sugar sequence $Man_3GlcNac_2$ and are always covalently attached to aspargine, specifically at the Asn-X-Ser/Thr protein sequons. They can be classified into three groups: (1) oligomannose – only mannose residues are attached to the core; (2) complex – two or more antennae are attached to core via N-acetylglucosamine (GlcNac); (3) hybrid – mannose residues attached to the Manα1-6 arm and one or two antennae to the Manα1-3 arm (Figure 2)[6]. An average N-glycan is a complex non-linear oligosaccharide composed of 10 to 15 monosaccharide residues. They are transferred to protein moiety on the luminal side of the endoplasmatic reticulum (ER) membrane while the protein itself is synthesized on ER-bound ribosomes and is translocating in the ER membrane.

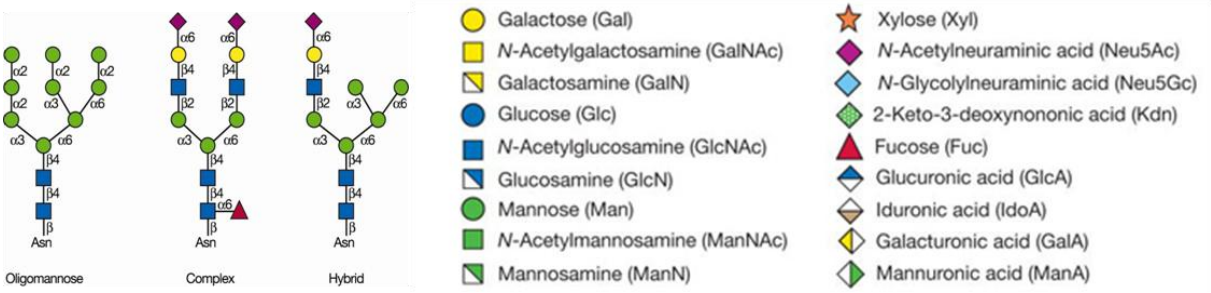


Figure 2: Symbolic representations of common monosaccharides and three general types of N-glycans, oligomannose, complex and hybrid. N-glycans are covalently attached to a peptide chain at aspargine and share the common core $Man_3GlcNac_2Asn$[6].

N-glycosylation is the most common posttranslational modification of plasma proteins. Deficient or defective protein N-glycosylation can lead to pathological conditions first of which was described in 1980, congenital disorder of glycosylation (CDG)[18]. Today more than 20 genetically caused disorders have been identified and the severity of these diseases demonstrates the importance of N-glycosylation[19].

One of the modifications in the process of the glycan maturation in vertebrate, which mostly occurs in the *trans*-Golgi, is the fucosylation (addition of the fucose at the

branched oligosaccharide), both core and antennary. Fucose can be added in an α1-6 linkage to the N-acetylglucosamine adjacent to aspargine in the core and in an α1-3 linkage as “capping” to the one or more branches (antennae) (Figure 3).

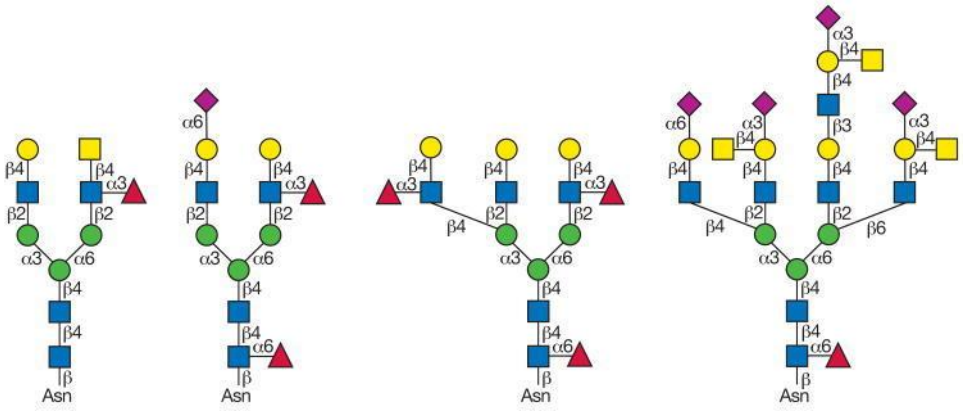


Figure 3: Structures of the complex fucosylated (red triangle) N-glycans[6].

Variations in the glycosylation of a single glycoprotein can have very profound and different consequences on its function. Absence of only one monosaccharide in the complex oligosaccharide structure can alter the function of the protein. The most prominent examples are sialylation and core fucosylation of the immunoglobulin G (IgG). IgG mediates pro- and anti-inflammatory activities through the engagement of its Fc fragment with distinct Fcγ receptors and those specific properties of IgG result from differential sialylation of the glycan attached to Fc part of the protein[20]. Another function changing sugar addition to the core of the glycan attached to the IgG is fucosylation. Increase of the fucosylation decreases antibody-dependent cell cytotoxicity (ADCC) and consequential tissue damage[21,22]. Those examples clearly demonstrate the importance of the oligosaccharide moiety of the proteins and what major significance the absence of only one monosaccharide from the glycan structure can have on the protein function.

1.3 N-GLYCAN ANALYSIS

Glycans can be analyzed from the body fluids or individual glycoproteins and the ability to capture and detect the diversified and heterogeneous array of glycans from the glycome is highly dependent on the availability of robust, high-resolution and sensitive separation techniques. The most important techniques used for the analysis of glycans are listed in (Table 1).

Table 1: The most commonly used techniques for the glycan analysis[6]

Acronym	Technique	Description	Use
FACE	fluorophore-assisted carbohydrate electrophoresis	gel-electrophoresis-based chromatographic technique for separating samples derivatized with an anionic fluorophore	separation, identification, and quantification of labeled mono- and oligosaccharides
GLC or GC	gas-liquid chromatography or gas chromatography	gas-phase chromatographic technique for separating volatile derivatized samples	sugar composition and linkage analysis; usually interfaced with MS
HPAEC--PAD	high-pH anion-exchange chromatography–pulsed amperometric detection	ion-exchange liquid chromatographic separation technique carried out at high pH	separation, identification, and quantification of mono- and oligosaccharides without derivatization
HPCE	high-performance capillary electrophoresis	chromatographic technique for separating charged molecules	separation, identification, and quantification of charged glycans; sometimes inter faced with MS
HPLC	high-pressure liquid chromatography	chromatographic technique for analytical and preparative separations	separation of all classes of glycans and glycoconjugates; may be interfaced with MS
SDS-PAGE	sodium dodecyl sulfate–polyacrylamide gel electrophoresis	gel electrophoresis technique for separation of proteins according to molecular weight	glycoprotein characterization
NMR	nuclear magnetic resonance	1D NMR spectroscopy	number and anomeric configuration of monosaccharides in a glycan
MS	mass spectrometry	technique for mass measurement of gas-phase ions	primary structure analysis of biopolymers
MALDI	matrix-assisted laser desorption ionization	MS ionization technique	mass mapping of glycans and glycoconjugates; important for glycomics
ESI	electrospray ionization	MS ionization technique	molecular weight and sequence analysis of glycans and glycoconjugates; glycoproteomics

1.3.1 Mass spectrometry (MS) analysis

Most commonly mass spectrometry methods used for oligosaccharide analysis are matrix-assisted laser desorption ionization (MALDI) and electrospray ionization (ESI)[23]. MS technologies allow the analysis of both, intact glycoconjugates and fragments thereof. Features that can be determined via MS are: degree of heterogeneity, type of glycosylation, site of glycosylation, branching pattern, number and lengths of branches, sequence of glycans, substitution with fucose, sialic acid and other capping groups[6]. Ionization behavior of glycoproteins is strongly influenced by the chemistry of the oligosaccharide moiety[24] by the increase of acidity, hydrophilicity and surface activity of the glycosylated proteins and peptides.

Sample preparation for MALDI mass spectrometry includes mixing the analyte with an acidic matrix, drying, and finally forming crystals. Pulsed laser light targeted at the crystallized glycans causes molecular plume containing ionized analyte and matrix molecules. Since acidic glycans monosaccharide dissociate to a large extent during the MALDI process, the loss of sialic acid must be taken into account when using this MS method.

Glycans analyzed by ESI enter the electroconductive needle in a stream of liquid. To the analyte solution, an elevated electrical potential of positive or negative polarity is applied that causes formation of extremely fine droplets that readily evaporate to form analyte ions. The electrospray process applies less vibrational energy on the glycan molecules with the result that fragmentation of glycoconjugates is generally not observed[6]. This method is particularly useful when complex mixtures peptides and glycopeptides are being analyzed.

1.3.2 High throughput quantitative HPLC analysis

Generally accepted method for quantitation of reductively animated glycans is chromatography. Glycans released from the glycoprotein are labeled with fluorescent dye, usually 2-aminobenzamide, 2-aminopyridine or 2-aminobenzoate which enables femtomole levels of oligosaccharide detection[25]. Hydrophilic interaction chromatography (HILIC), a variant of the normal-phase chromatography, is the method of choice for the glycan analysis. The stationary phase retains a semi-immobilized layer of mostly organic mobile phase (70-80% of acetonitrile) enriched with water. Glycans are retained on the NH_2-silica stationary phase because of their

hydrophilic basic groups what represents the main mechanism of HILIC separation of oligosaccharide: partitioning between stagnant aqueous layer and the bulk of the organic phase[26]. The mobile phase is the mixture of organic solution, usually acetonitrile, and polar solvent, ammonium formate for example. Gradient elution of the glycans is achieved by gradually increasing the polarity of the mobile phase by reducing the percentage of organic phase and proportionally increasing the salt gradient. In this way larger oligosaccharides, which are more hydrophilic, require a higher percentage of polar mobile phase and more time to elute.

HILIC analysis offers several advantages over other commonly used techniques. Firstly, it is a reliable quantitative method proven to provide quality data for the complex studies which use large cohorts of samples[27]. Secondly, charged and neutral glycans can be analyzed in the same HPLC run in contrast to capillary electrophoresis (CE), where sialic acids have to be removed because glycans migration through capillary is dependent only on charged covalently attached label[28], and in contrast to mass spectrometry, where sialic acid linkages are unstable[29]. Thirdly, glycans with same composition, but different linkage or sequence, can be efficiently separated as in the case where a glycan with a core α1-6 fucose monosaccharide elutes earlier than one with the fucose α1-3 linked to an outer arm. And finally, the none selective efficient 2AB labeling gives 1:1 stoichiometry allowing accurate quantitative measurement of the relative amounts of individual glycans[30].

One of the approaches which enabled a huge breakthrough in quantitative plasma N-glycan analysis is high throughput quantitative HPLC analysis based on a 96-well plate format which enables release, purification and labeling of glycans from 96 samples in the period of 3 days[30] (Figure 4).

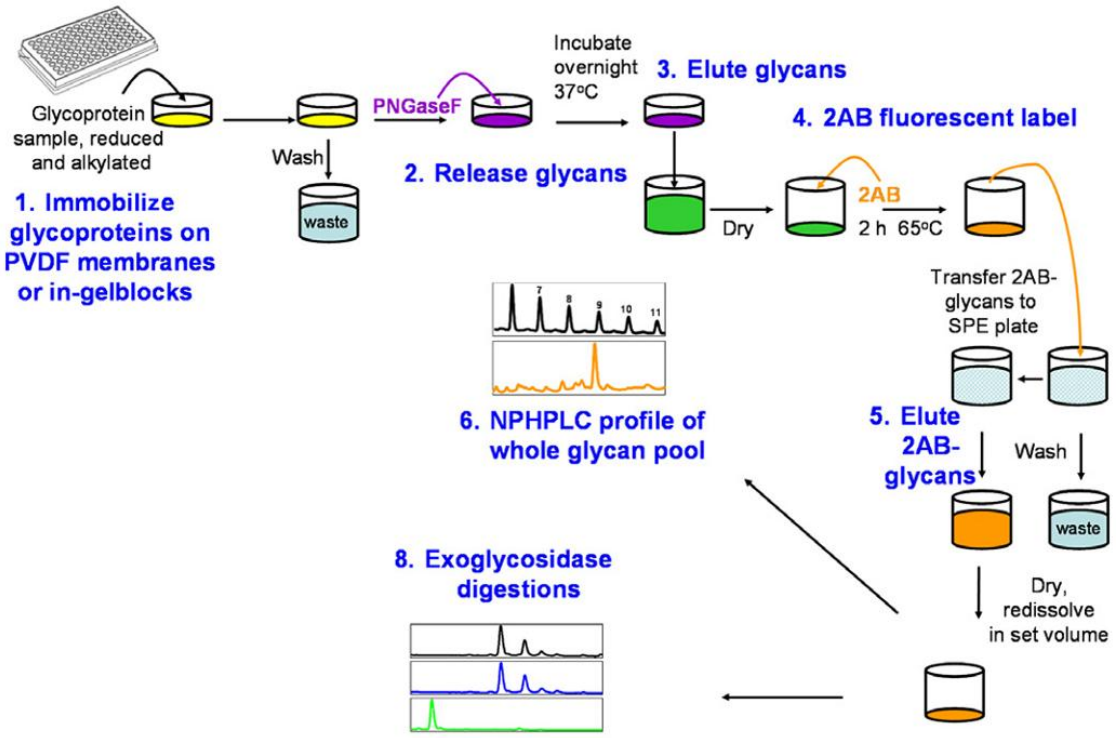


Figure 4: Glycan release, fluorescent labeling, HILIC profiling, structure assignments, and relative quantitation[30].

Such advances in high-throughput glycan analysis allowed reliable quantification of N-glycans in large cohorts[9,31] which lead to fruitful research of the genetic regulation and biological roles of glycan structures and brought glycomics into line with genomics, proteomics and metabolomics[32].

1.4 Genome wide association study (GWAS)

Genome-wide association study (GWAS) is technology that analyzes DNA sequence variation from across the human genome in order to identify genetic polymorphisms accounting for different phenotypes. This kind of studies are primarily preformed for the purpose of identifying genetic risk factors for common complex diseases such as schizophrenia and diabetes in order to be able to predict who is at risk and to identify biological causation of the deseases[33]. Identification of the *complement factor H* gene as a major risk factor for age related macular degeneration was one of the first major successes of this approach[34,35]. Analyzed unity of genetic variations is single nucleotide polymorphism (SNP) which is defined as single base-pair changes in the DNA that occur with high frequency in the human genome[36]. SNPs are the most abundant form of genetic variation in the genome and are therefore used as markers

of the genetic risk factor located in the genome. They themselves can, but usually do not, have functional consequences, for example amino acid changes, changes to mRNA transcript stability and changes to transcription factor binding affinity[37].

In accordance to common disease/common variant (CD/CV) hypothesis[38], the total genetic risk due to common genetic variation must be spread across multiple genetic factors. Glycosylation, as already mentioned, is a complex biosynthetic pathway which is not template driven but rather represents a highly complicated and dynamic network of hundreds of genes coding for enzymes and other proteins involved in the synthesis of glycans. In that sense, glycosylation variability, due to common genetic variation, must also be spread across multiple genetic factors. This suggest that traditional family-based genetic studies are not likely to be successful for determining the genetic portion of glycosylation variability as well, hence the studies have turned towards large population cohorts[27,39]. This approach proves to be fruitful and led, among others, to the results presented in this thesis.

1.5 Glycans as biomarkers

Glycans have only recently been introduced as potential disease biomarkers. Progress has been hindered by the lack of competent analytical methods. Alterations in glycosylation are associated with a wide range of diseases and a number of different glycan biomarkers have been identified[40–42]. The potential of glycan biomarkers stems from the fact that there is no direct genetic template for glycans. While the protein sequence is genetically encoded with a direct correlation between nucleotide and amino acid sequence, such templates for glycans do not exist. Hence, glycan biosynthesis is more significantly affected by disease states than protein production. Both genetic polymorphisms and environmental effects affect the performance of the glycosylation pathway. Glycans are essentially metabolic products that are amplified with even small changes in the associated protein (glycosyltransferases) expression, hence the effect of the disease is significantly amplified and easier to detect[40]. Glycosylation processing pathways are cell type and protein specific. Therefore differences in glycan processing steps are sometimes specific for a given glycoprotein, or a subset of glycoproteins, while in other cases they may be more general and affect many glycoproteins. This peculiarity of the glycosylation process opens the possibility of revealing functionally important

changes in the glycosylation of proteins that are confined to inaccessible tissues by analyzing easily accessible plasma glycoproteins.

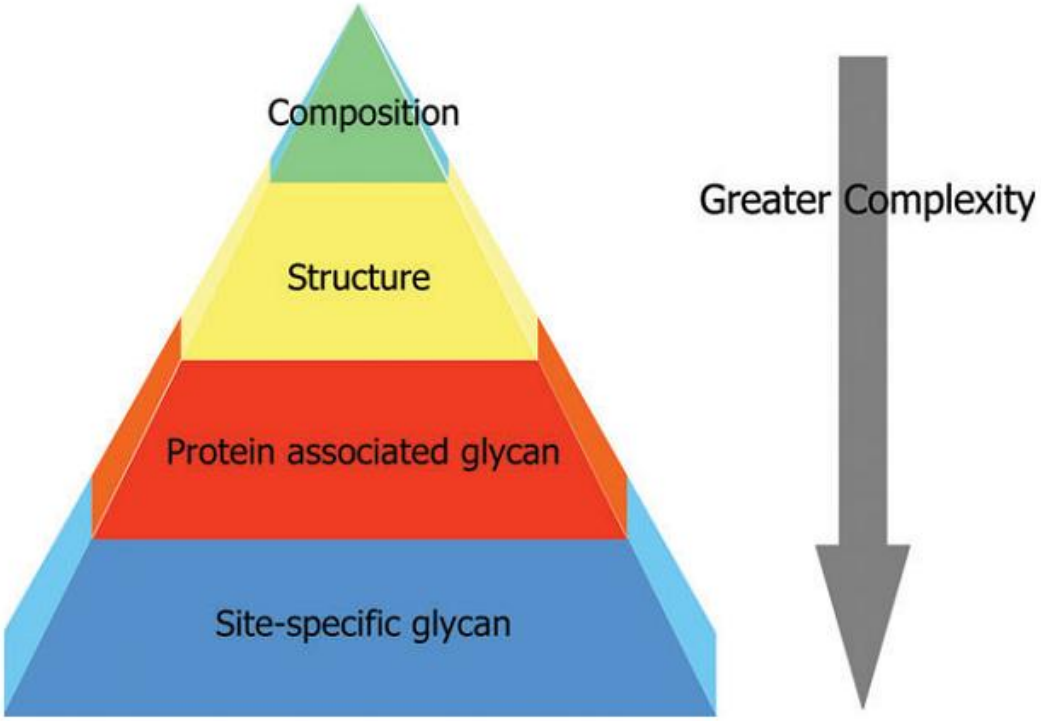


Figure 5: Complexity levels of oligosaccharide analysis. Glycans have biomarker potential at the level of simple composition analysis (increased / decreased fucose, sialic acid etc.) up to the level of complex analysis of a single protein site specific glycosylation[40].

Glycans show biomarker potential at several levels of complexity. Even at the level of easily detectable simple changes in the composition of the plasma glycome, like the decreased fucosylation of increased sialylation, can provide valuable information about the onset or progress of the disease (Figure 5). It is not necessary to identify the exact glycan structure or a protein it is attached to. On the other hand, protein specific or even site specific glycosylation offers a rich, valuable and informative source of glycan markers. It is predict that, in the future, it will be possible to not only diagnose diseases, but perhaps even determine disease progression and specific strain based solely on glycan profiling[40].

1.6 MATURITY ONSET DIABETES OF THE YOUNG (MODY)

Maturity onset diabetes of the young represents a heterogeneous group of disorders characterized by diabetes mellitus. From a clinical point of view it is defined as autosomal dominantly inherited, non-insulin-dependent, early onset diabetes. MODY can result from mutation in several genes and by now at least eight genetic subgroups of MODY are known from which mutations in glucokinase gene (GCK) and transcription factors (mostly hepatic nuclear factors - HNF) cover 88% of cases (Figure 6)[43].

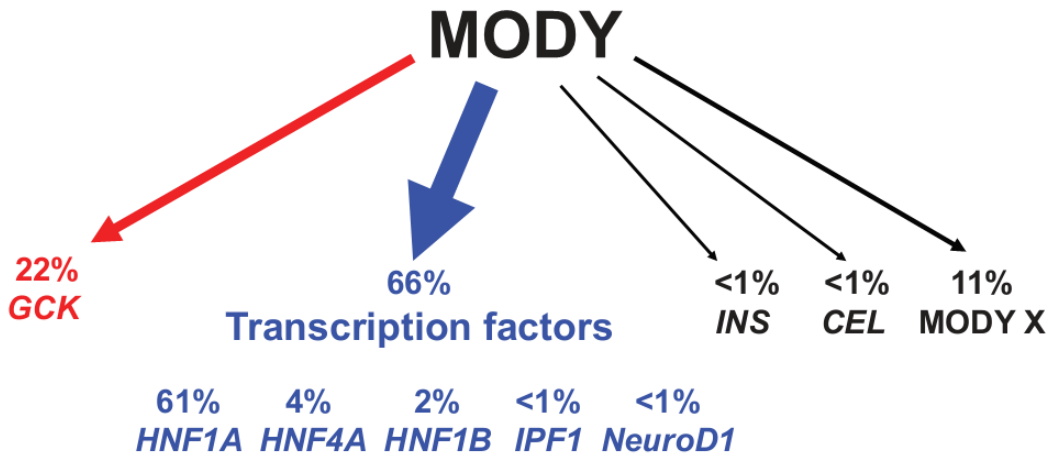


Figure 6: Genetic heterogeneity of clinically defined maturity onset diabetes of the young[43].

All causative genes are expressed in the pancreas β cells and mutation in any of them leads to abnormal β cell function and finally to diabetes mellitus (Figure 7). Glucokinase (GCK associated with MODY2[44,45]) is enzyme that catalyses the transfer of phosphate from ATP to glucose to form glucose-6-phosphate which enters the glycolysis. ATP released during glycolysis and Krebs cycle leads to the closure of the ATP-sensitive potassium channel, depolarization of the cell membrane and opening of the voltage-dependent calcium channel. Following influx of calcium and mobilized intracellular calcium are causing the fusion of insulin-containing secretory granules with the cell membrane and finally the release of insulin into circulation[46]. Transcription factors associated with MODY (hepatocyte nuclear factor (HNF) 4 a (associated with MODY 1[47]), HNF-1 a (MODY 3[48]), insulin promoter factor 1 (IPF-1 [MODY 4][49]), HNF-1 b (MODY 5[50]), and neurogenic differentiation factor 1 (NeuroD1[51]) are being expressed in the nucleus of beta-cells and regulate the transcription of the insulin gene, transcription of genes encoding enzymes involved in

the transport and metabolism of glucose as well as other proteins required for normal function of β cells.

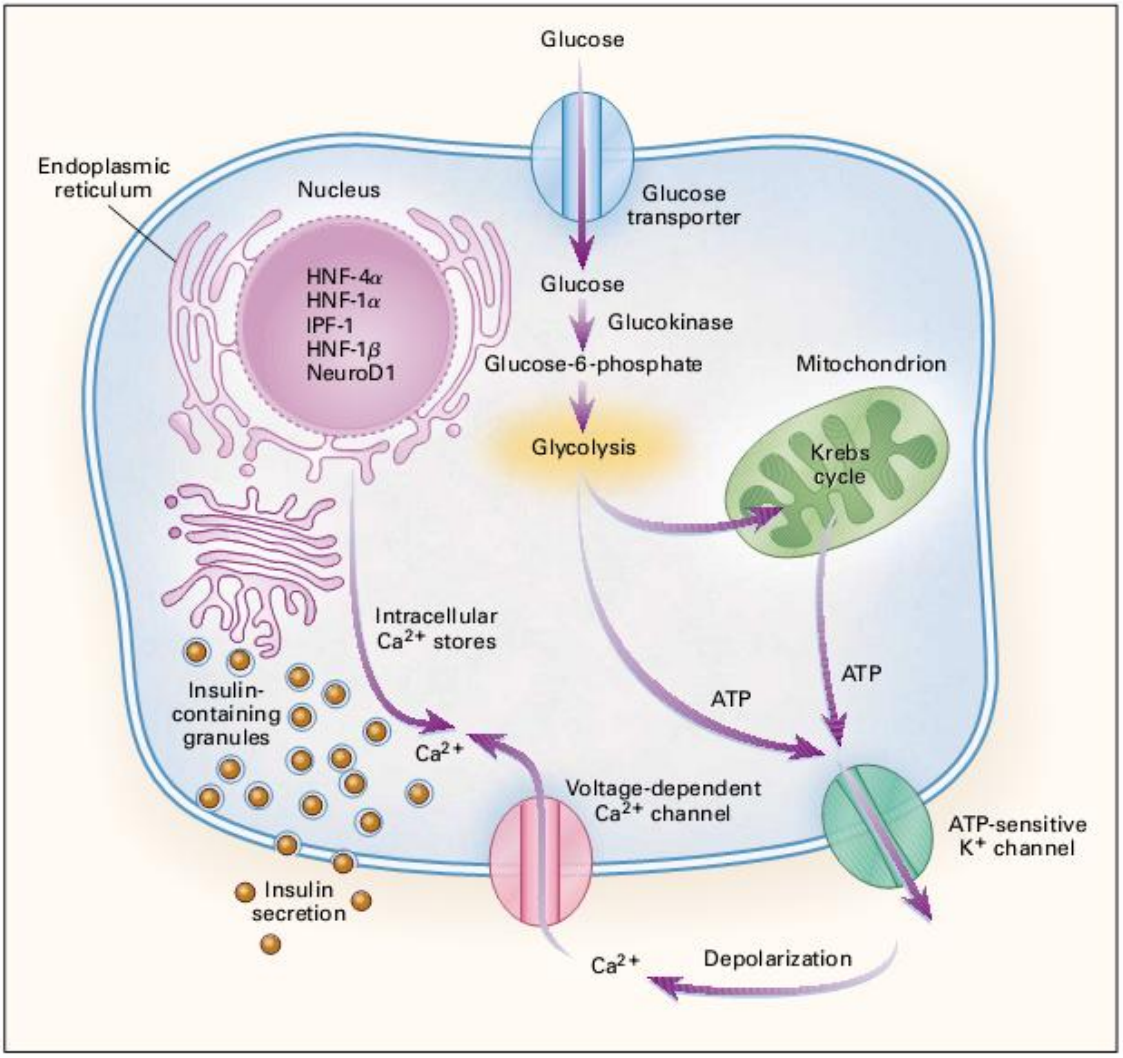


Figure 7: Enzymes, transporters and transcription factors implicated in the molecular mechanism of Maturity-Onset Diabetes of the Young[52].

1.7 HNF1A TRANSCRIPTION FACTOR AND HNF1A-MODY

HNF1A-MODY is caused by mutations within 10 exons, promoters and splice sites of transcription factor hepatocyte nuclear factor (HNF)-1A (Figure 8) that regulates gene expression in pancreatic islets, liver, kidney and intestine and is one of the key regulators of metabolic genes[53]. A total of 194 mutations have been reported in 373 families and they include missense, nonsense, insertions, duplications, deletions, promoter region mutations and splice sites mutations[54].

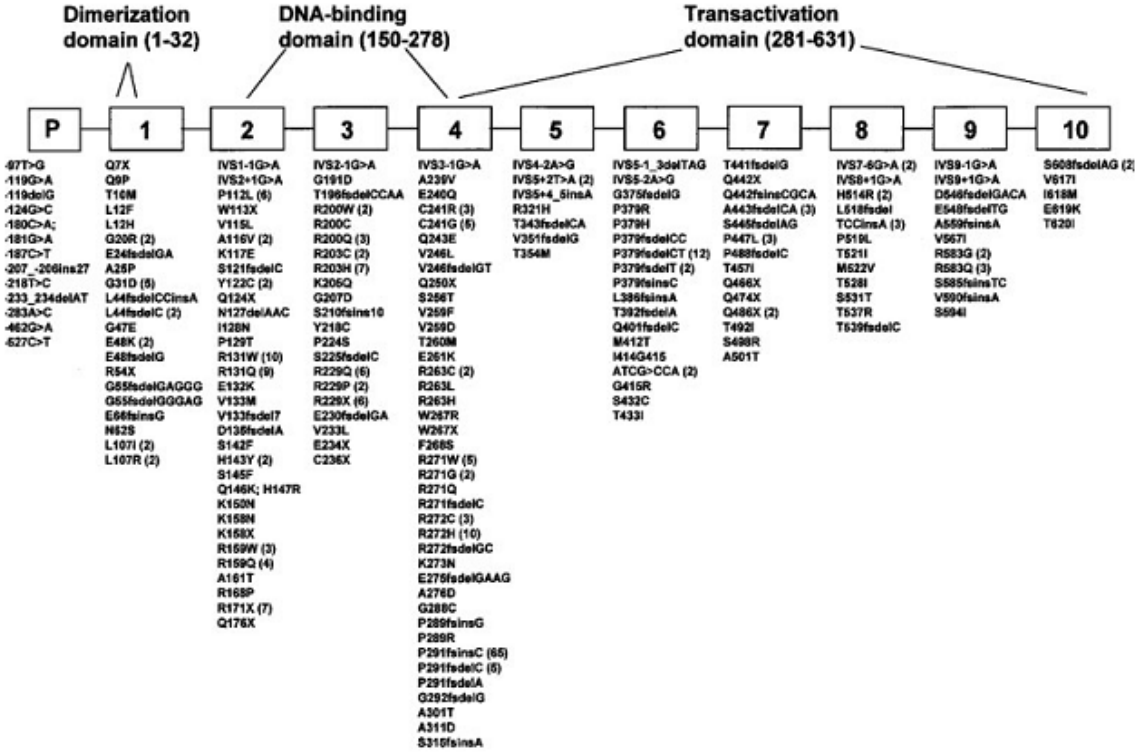


Figure 8: Map of human hepatocyte nuclear factor 1-alpha (HNF1A) mutations within 10 exons, promoters and splice sites. The numbers in brackets next to the mutation names refer to the number of families the mutations was reported in[54].

Patient carrying those mutations develop diabetes symptoms in teens or early adult life and are marked with progressive β-cell failure that leads to progressive increase of hyperglycemia through life. It is estimated that they account for approximately 1-2 % of diabetes with the population frequency of roughly 0.02-0.04 %. Due to the onset of symptoms in slim young adults, HNF1A-MODY is frequently misdiagnosed as type 1 or type 2 diabetes (Table 2)[55]. Since the therapy for this type of monogenic β-cell diabetes is different to those used to treat type 1 or type 2 diabetes, wrong diagnosis and treatment can significantly lower the quality of life of misdiagnosed patients. This fact highlights the importance of accurate and effective HNF1A-MODY diagnosis as those patients are very sensitive to sulfonylurea therapy (Figure 9)[56–59]. Sulfonylurea works by inhibiting the sulfonylurea receptor 1 (SUR1), the regulatory subunit of the ATP-sensitive potassium channels (K_{ATP}) in pancreatic cells allowing the membrane potential to rise. Elevated membrane potential causes the opening of voltage gated Ca^{2+} channels, enhanced Ca^{2+} entry, and finally secretion of insulin (Figure 10)[56].

Table 2: Differentiation of β-cell monogenic diabetes from type 1 and type 2 diabetes.
[a] The population frequency is the frequency of obesity that occurs in the general population. Abbreviations: 3243 MIDD, maternally inherited diabetes and deafness associated with mitochondrial m.3243A>G mutation; GCK-MODY, maturity onset diabetes of the young associated with mutations in glucokinase; K_{ATP}PNDM, permanent neonatal diabetes associated with mutations in the ATP-sensitive potassium channel; TF-MODY, maturity onset diabetes of the young associated with mutations in transcription factors (e.g. hepatocyte nuclear factor 1-α[HNF1-α], HNF4-α, or HNF-1β).[60]

Features	Type 1 diabetes	Young-onset type 2 diabetes	GCK-MODY	TF-MODY	K_{ATP}PNDM	3243 MIDD
Insulin dependence	yes	no	no	no	no	yes or no
Parent affected	2-4%	yes	yes	yes	15%	mother
Age of onset	6 months to adulthood	adolescence and young adulthood	birth	teens to young adulthood	< 6 months	young adulthood
Obesity[a]	population frequency	increased frequency	population frequency	population frequency	population frequency	rare
Acanthosis nigricans	no	yes	no	no	no	no
Glycemia	high	variable	mild	high	high	variable
β-Cell autoantibodies	yes	no	no	no	no	no
C-peptide (nmol/L)	<0.33	0.5->1	0.1-0.7	0.1-0.7	<0.2	0.1-0.7

HNF1A-MODY patients show no insulin secretory response to intravenous glucose in contrast to a strong response to intravenous sulfonylurea what indicates the nature of β-cell defect[56]. Hnf-1α$^{-/-}$ mouse islets displayed a reduction in mRNA levels of proteins involved in glucose uptake and glycolysis[61,62] and mitochondrial metabolism[61] what results in decreased ATP production, impaired glucose metabolism, and decreased insulin release[63]. Glycolysis and mitochondrial ATP production are upstream of the SUR1 what enables sulphonylurea to bypass all major β-cell defects caused by mutations in HNF1A gene[56].

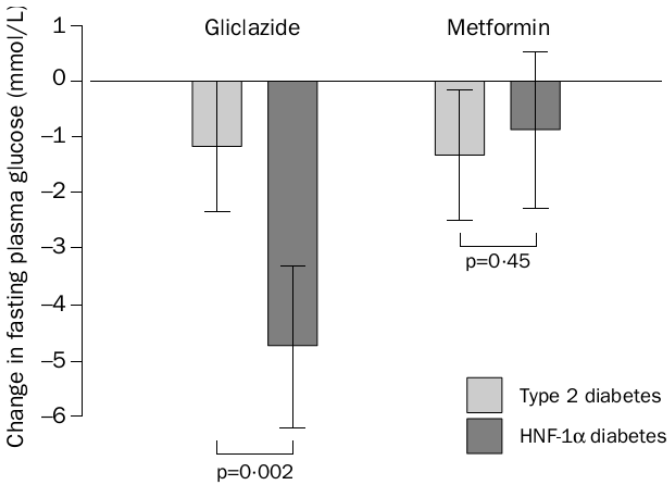


Figure 9: Response to sulfonylurea (gliclazide) and metformin in type 2 diabetes and HNF1A MODY[56].

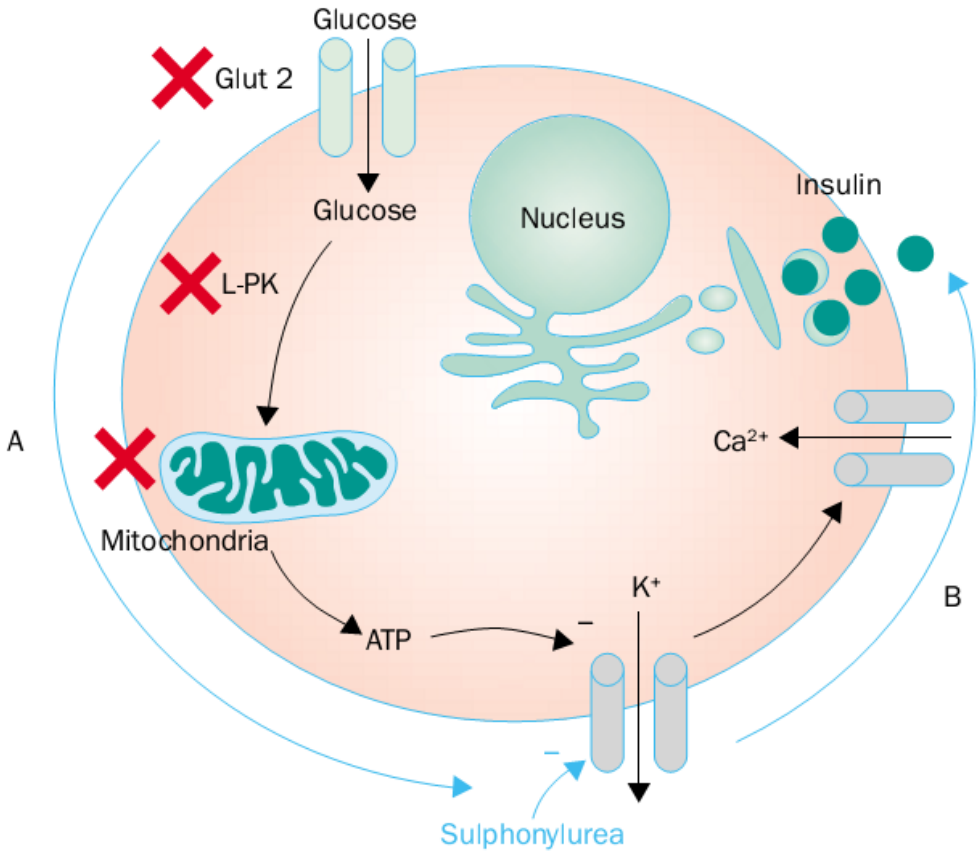


Figure 10: Sulphonylurea circumvents severe defects in glucose-induced insulin secretion of HNF1A MODY patients and stimulates insulin secretion by binding to SUR1 subunit of the ATP-sensitive K^+ channel. L-PK=L-type pyruvate kinase; A=pre-K_{ATP} channel; B=post-K_{ATP} channel[56].

2 RESEARCH OBJECTIVES

In the genome wide association study (GWAS) several single nucleotide polymorphisms (SNPs) in the region of HNF1A gene displayed significant association with glycosylation traits. Objective of this work is to explore the role of transcription factor HNF1A, previously not associated with N-glycans, in the regulation of protein N-glycosylation.

Plasma proteins N-glycosylation of HNF1A-MODY (maturity onset diabetes of the young) patients, carrying inactivation mutations in HNF1A transcription factor, will be compared to glycosylation of general population and four cohorts diagnosed with different diabetes types.Potential of plasma protein glycans as specific biomarker for HNF1A-MODY will be evaluated

Robustness of the high throughput platform used for the glycan analysis is going to be tested in order to improve the quality of obtained results and to minimize the variability introduced by the method.

3 MATERIALS AND METHODS

3.1 STUDY OF GENERAL POPULATIONS

Three populations of adult individuals were recruited within a community irrespective of any specific phenotype: the CROATIA-VIS, CROATIA-KORCULA (both cohorts from the Croatian Dalmatian islands recruited in 2003–2004 and 2007 respectively), and the Orcades population which is part of ongoing study with participants recruited from the Orkney islands in Scotland. Fasting blood samples were collected, biochemical and physiological measurements taken and questionnaires of medical history as well as lifestyle and environmental exposures collected following similar protocols. The CROATIA-VIS study includes 1008 Croatians, aged 18 and 93 years, who were recruited from the villages of Vis and Komiža on the Dalmatian island of Vis during 2003 and 2004 within a larger genetic epidemiology program[64]. The CROATIA-KORCULA study includes 969 Croatians between the ages of 18 and 98[64]. The field work was performed in 2007 in the eastern part of the island, targeting healthy volunteers from the town of Korčula and the villages of Lumbarda, Žrnovo and Račišće. The Orkney Complex Disease Study (ORCADES) is an ongoing study in the isolated Scottish archipelago of Orkney[64]. Data for participants aged 18 to 100 years, from a subgroup of ten islands, were used for this analysis. DNA samples were genotyped according to the manufacturer's instructions on Illumina Infinium SNP bead microarrays (Hu-manHap300v1 for the CROATIA-VIS cohort, HumanHap300v2 for the ORCADES cohort and HumanCNV370v1 for the CROATIA-KORCULA cohort). Genotypes were determined using Illumina BeadStudio software. Genotyping was successfully completed on 991 individuals from CROATIA-VIS, 953 from CROATIA-KORCULA and 761 from ORCADES.

3.2 DIABETES SUBTYPES STUDY POPULATIONS

3.2.1 Subjects from Norway

The MODY subjects were recruited through the Norwegian MODY Registry, comprising MODY patients from all over Norway. The MODY subjects have mutations (confirmed by a certified Norwegian diagnostic centre) in HNF1A (n=47 from 5 families), GCK (n=30 from 19 families) or HNF4A (n=10 from 4 families). Type 2 diabetes patients (n=16) were diagnosed ≤45 years and recruited at the diabetes outpatient clinic at Haukeland University Hospital. All subjects were of North European ethnicity.

3.2.2 Subjects from Slovakia

Subjects fulfilling the MODY clinical diagnostic criteria were actively searched in diabetes outpatient clinics throughout Slovakia. Patients with fasting hyperglycemia were tested for GCK mutations first and if negative, then for HNF1A and HNF4A mutations; patients with higher HbA1c and requiring insulin treatment were tested for HNF1A mutations first and then if negative, were tested for HNF4A and GCK mutations. All patients from Slovakia carrying HNF1A (n=46), HNF4A (n=2) or GCK (n=77) mutations with relevant biochemical and clinical data were included in this study. All subjects were white Europeans.

3.2.3 Subjects from Oxford, UK

Subjects were ascertained from the South of England. The MODY samples comprise subjects with a mutation (confirmed by sequencing in a certified UK diagnostic centre) in either HNF1A (n=34) or HNF4A (n=5). Type 2 diabetes subjects were selected from the Young Diabetes in Oxford (YDX) study, comprising subjects diagnosed with diabetes ≤45 years of age. Criteria for diagnosis were: C-peptide positive, no requirement for permanent insulin within 3 months of diagnosis and negative GAD antibodies. Type 1 diabetes subjects were selected from the YDX study and criteria for diagnosis were: permanent insulin therapy from diagnosis and C-peptide<0.1nmol/l and/or positive glutamic acid decarboxylase antibodies. Subjects with either a clinical label of type 1 diabetes or type 2 diabetes did not meet clinical criteria for MODY diagnostic testing or had been tested and were negative for mutations in HNF1A, HNF4A or GCK. The Oxford samples were all plasma samples. The blood was taken from all examinees at the time of recruitment and their plasma or serum samples were stored at -20°C until the analysis.

3.3 ETHICS STATEMENT

All studies conformed to the ethical guidelines of the 1975 Declaration of Helsinki and were approved by appropriate ethics boards with all respondents signing informed consent prior to participation.

3.4 GLYCAN RELEASE AND LABELING

Plasma (or serum) samples (5 µL) were reduced in a polypropylene 96-well flat-bottomed microplate by adding 2 µL of 5 x sample buffer (0.625 mL of 0.5 M Tris [pH 6.6], 1 mL of 10% SDS, and 3.375 mL of water), 2 µL of water, and 1 µL of 0.5 M dithiothreitol (DTT) and then were incubated at 65 °C for 15 min. The samples were then alkylated by adding 1 µL of 100 mM iodoacetamide and were incubated for 30 min in the dark at room temperature. The samples were then set into a gel block by adding 22.5 µL of 30% (w/w) acrylamide/0.8% (w/v) bis-acrylamide stock solution (37.5:1.0, Protogel, National Diagnostics, Hessle, Hull, UK), 11.25 µL of 1.5 M Tris (pH 8.8), 1 µL of 10% SDS, 1 µL of 10% ammonium peroxodisulfate (APS), and finally 1 µL of N,N,N,N´-tetramethyl-ethylenediamine (TEMED), mixed, and then left to set. Formed gel blocks were transferred to a filter plate (Whatman protein precipitation plate) and then washed with 1 mL of acetonitrile with vortexing on a plate mixer (Gesellschaft Labortechnik (GFL) GFL-3023, Germany) for 10 min, followed by removal of the liquid (all washes and elutions were per- formed on a vacuum manifold). This washing procedure was repeated twice with 1 ml of 20 mM $NaHCO_3$ (pH 7.2) followed by 1 mL of acetonitrile. N-glycans were released by adding 50 µL of 0.1 U/mL PNGaseF (Prozyme, Leandro, CA, USA) in 20 mM $NaHCO_3$ (pH 7.2) to reswell the gel pieces for 5 min, 50 µL of 20 mM $NaHCO_3$ (pH 7. 2) was added, and then the plates were sealed with adhesive film (Peelable adhesive seals, Mettler Toledo) and incubated overnight at 37 °C. The released glycans were collected in a 2 mL polypropylene 96-well plate by washing the gel pieces with 3 x 200 µL of water, 200 µL of acetonitrile, 200 µL of water, and finally 200 µL of acetonitrile. The released glycans were dried, 20 µL of 1% formic acid was added, and the mixture was incubated at room temperature for 40 min and then redried. Samples for HPLC analysis were labeled by adding 5 µL of 2AB labeling solution (55 mg Anthranilamide (> 98%, Sigma) + 66 mg Sodium Cyanoborohydride (95%, Sigma) + 330 µl Glacial Acetic Acid (Merck) + 770 µl DMSO (Sigma)), vortexed for 10 min, incubated for 30 min at 65 °C, vortexed again for 10 min, and incubated for a further 90 min. Excess 2AB was removed using solid-phase extraction (SPE) with Whatman 3MM chromatography paper. For paper cleanup, 1-cm square pieces of prewashed, dried Whatman 3MM chromatography paper were folded into quarters and placed into a filter plate (Whatman protein precipitation plate prewashed with 200

µL of acetonitrile followed by 200 µL of water). The 5 µL of 2AB-labeled samples were applied to the paper and left to dry/bind for 15 min. The excess 2AB was washed off the paper by vortexing with 2 mL of acetonitrile for 15 min and then removing the acetonitrile using a vacuum manifold; this procedure was repeated four times. The labeled glycans were eluted from the paper by vortexing with 900 µL of water for 30 min and then collected by vacuum into a 2-ml 96-well plate. This was repeated with a further 900 µL of water. The labeled glycans were then eluted into a 2-ml 96-well plate with 400 µL of 200 mM ammonia solution under gravity for 20 to 30 min before a low vacuum was applied to elute the last drops. The eluted 2AB-labeled glycans were dried before redissolving in a known volume of water ready for analysis by HILIC.

3.5 SIALIDASE DIGESTION

HPLC quantification sialidase digestion was performed to improve measurement precision. Aliquots of the 2-AB-labeled glycan pool were dried down in 200-mL microcentrifuge tubes. To these, the following was added: 1 ml of 500 mM sodium acetate incubation buffer (pH 5.5), 1ml (0.005 units) of ABS, Arthrobacter ureafaciens sialidase (releases a2–3, 6, 8 sialic acid, Prozyme) and H_2O to make up to 10mL. This was incubated overnight (16–18 h) at 37 °C and then passed through a Micropure-EZ enzyme remover (Millipore, Billerica, MA, USA) before applying to the HPLC.

3.6 HYDROPHILIC INTERACTION HIGH-PERFORMANCE LIQUID CHROMATOGRAPHY

Released glycans were subjected to hydrophilic interaction high performance liquid chromatography (HILIC) on a 25064.6 mm i.d. 5 mm particle packed TSKgel Amide 80 column (Tosoh Bioscience, Stuttgart, Germany) at 30 °C with 50 mM formic acid adjusted to pH 4.4 with ammonia solution as solvent A and acetonitrile as solvent B. 60 min runs were on a 2795 Alliance separations module (Waters, Milford, MA). HPLCs were equipped with a Waters temperature control module and a Waters 2475 fluorescence detector set with excitation and emission wavelengths of 330 and 420 nm, respectively. The system was calibrated using an external standard of hydrolyzed and 2-AB-labeled glucose oligomers from which the retention times for

the individual glycans were converted to glucose units (GU)[65]. Glycans were analyzed on the basis of their elution positions and measured in glucose units then compared to reference values in NIBRT's ''GlycoBase v3.0 '' database available at http://glycobase.nibrt.ie) for structure assignment[65].

3.7 GLYCAN STRUCTURAL FEATURES

Levels of glycans sharing the same structural features were approximated by adding the structures having same characteristic, from either HILIC, or after sialidase treatment integrated glycan profiles (individual glycan structures present in each glycan group reported previously are shown in Table 3 and Table 4). Glycan features were defined as: Core fucosylated glycans (FUC-C) = DG6/(DG5+DG6)*100; Antennary fucosylated glycans (FUC-A) = DG7/(DG5+DG7)*100; Biantennary glycans (BA) = DG1+DG2+DG3+DG4+DG5+DG6+DG7; Monosialylated biantennary glycans (BAMS) =(G7+G8)/(DG5+DG6+DG7)*100; Disialylated biantennary glycans BADS=(G9+G10+G11)/(DG5+DG6+DG7)*100; Triantennary glycans (TRIA) = DG8+DG9+DG10; Tetraantennary glycans (TA) = DG11+DG12+DG13; Nongalactosylated glycans (G0) = DG1+DG2; Monogalactosylated glycans (G1) = DG3+DG4; Digalactosylated glycans (G2) = DG5+DG6+DG7; Trigalactosylated glycans (G3) = GP12+GP13+GP14; Tetragalactosylated glycans (G4) = GP15+GP16.

Table 3: Glycan structures present in 16 initial profile HPLC peaks

Peak	Structure	Peak	Structure	Peak	Structure
GP1	A2	**GP7**	FA2BG2 M7D3 A2G2S(3)1 A2G2S(6)1 M7D1	**GP12**	A2F1G2S2 A3G3S(3.3)2 A3G3S(3.6)2 A3G3S(6.6)2 A3BG3S(3.3)2 A3BG3S(3.6)2 A3BG3S(6.6)2
GP2	A2B A1G1 FA2				
GP3	M5 FA2B A2[6]G1 A2[6]BG1	**GP8**	A2BG2S(3)1 A2BG2S(6)1 M5A1G1S1 FA2G2S(3)1 FA2G2S(6)1 A3G3 FA2BG2S(3)1 FA2BG2S(6)1	**GP13**	A3F1G3S2 FA3G3S(3.3)2 FA3G3S(3.6)2 FA3G3S(6.6)2 FA3BG3S(3.3)2 FA3BG3S(3.6)2 FA3BG3S(6.6)2 A3G3S(3.3.6)3 A3G3S(3.6.6)3 A3G3S(6.6.6)3
GP4	A2[3]G1 A2[3]BG1 M4A1G1 FA2[6]G1 FA2[6]BG1 A1[6]G1S(3)1 A1[6]G1S(6)1 FA2[3]G1 FA2[3]BG1 M6D1. D2 A1[3]G1S(3)1 A1[3]G1S(6)1	**GP9**	A2F1G2S(3)1 A2F1G2S(6)1 M8D2. D3 A2G2S(3.3)2 A2G2S(3.6)2 A2G2S(6.6)2 M8D1.D3	**GP14**	A3F1G3S(3.3.6)3 FA3G3F1S(6.6.6)3 A4G4S(6.6)2 A3F1G3S(3.6.6)3 A3F1G3S(6.6.6)3 A4G4S(6.6.6)3 A4F1G4S2 A4G4S3
GP5	M6D3 A2[6]G1S(3)1 A2[6]G1S(6)1 A2G2 A2[3]G1S(3)1 A2[3]G1S(6)1 A2BG2	**GP10**	A2BG2S(3.3)2 A2BG2S(3.6)2 A2BG2S(6.6)2 A3BG3S(3)1 A3BG3S(6)1 FA2G2S(3.3)2 FA2G2S(3.6)2 FA2G2S(6.6)2	**GP15**	A4G4S4 A4F1G4S3
GP6	FA2[6]G1S(3)1 FA2[6]G1S(6)1 FA2[6]BG1S(3)1 FA2[6]BG1S(6)1 M4A1G1S1 FA2G2 FA2[3]G1S(3)1 FA2[3]G1S(6)1 A2BG1S1 FA2[3]BG1S(3)1 FA2[3]BG1S(6)1	**GP11**	FA2BG2S(3.3)2 FA2BG2S(3.6)2 FA2BG2S(6.6)2 M9	**GP16**	A4G4S(6.6.6.6)4 A4G4S(3.6.6.6)4 A4BG4S4 FA4G4S4 A4F1G4S4 A4G4LacS4 A4F2G4S4 FA4F1G4S4

Table 4: Glycan structures present in 13 HPLC peaks after sialidase digestion.

Plasma N-glycans after sialidase digestion			
Glycan peak	**Peak composition**	**Glycan peak**	**Peak composition**
DG1	A2	**DG7**	M7D3
DG2	A2B		A2F1G2
	A1G1		M7D1
	FA2	**DG8**	A3G3
DG3	M5		A2F2G2
	FA2B		FA3G3
	A2[6]G1		M8D2, D3
	A2[6]BG1		M8D1,D3
DG4	M4A1G1	**DG9**	FA3BG3
	A2[3]G1		A3F1G3
	A2[3]BG1	**DG10**	M9
	FA2[6]BG1		FA3F1G3
	FA2[3]G1	**DG11**	A4G4
	FA2[3]BG1		A4BG4
DG5	M6D1, D2		A3F2G3
	M6D3		FA4G4
	A2G2	**DG12**	A4F1G4
	A2BG2	**DG13**	A4G4Lac
DG6	FA2G2		A4F2G4
	M5A1G1		FA4F1G4
	FA2BG2		

3.8 DESIGN OF EXPERIMENTS.

Placket-Burman (PB) screening designs are used to observe n-1 variables in n experiments proposing experimental designs for more than seven factors, and especially for n x 4 experiments, i.e. 8, 12, 16, 20 etc., suitable for the studying up to 7, 11, 15, 19 etc. factors, respectively. PB designs are two level designs in which every factor is observed at high, denoted as 1 and low level, denoted as -1[66]. In the present study the 11-factor, 12 experiment design was chosen as shown in Table 5.

Table 5: Plackett-Burman design in 12 runs (R1 – R12) for 11 factors. Upper level (+1) is labeled as +, and lower level (-1) as -.

Runs	Factors										
	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11
R1	+	+	+	+	+	+	+	+	+	+	+
R2	-	+	-	+	+	+	-	-	-	+	-
R3	-	-	+	-	+	+	+	-	-	-	+
R4	+	-	-	+	-	+	+	+	-	-	-
R5	-	+	-	-	+	-	+	+	+	-	-
R6	-	-	+	-	-	+	-	+	+	+	-
R7	-	-	-	+	-	-	+	-	+	+	+
R8	+	-	-	-	+	-	-	+	-	+	+
R9	+	+	-	-	-	+	-	-	+	-	+
R10	+	+	+	-	-	-	+	-	-	+	-
R11	-	+	+	+	-	-	-	+	-	-	+
R12	+	-	+	+	+	-	-	-	+	-	-

X1: duration of the exposure of proteins to the reducing agent; X2: duration of the exposure to the alkylating agent iodoacetamide; X3: whether the polyacrylamide gel blocks comprising proteins should be freezed; X4: concentration of N-glycosidase-F; X5: duration of the enzyme digestion; X6: storage of the eluted and dried glycans after the enzymatic digestion; X7: duration of the formic acid treatment; X8: duration of the 2-AB binding reaction; X9: time of the glycan binding to the chromatography paper; X10: duration of acetonitrile wash; X11: the storage time prior to the HPLC analysis.

Factors for screening were selected after identifying potentially critical steps in the three day method: duration of the exposure of proteins to the reducing agent

dithiotreitol (X1), duration of the exposure to the alkylating agens iodoacetamide (X2), whether the polyacrylamide gel blocks comprising proteins should be freezed (X3), concentration of N-glycosidase-F used for the glycan release (X4), duration of the enzyme digestion (X5), storage of the eluted and dried glycans after the enzymatic digestion (X6), duration of the formic acid treatment (X7), duration of the 2-AB binding reaction (X8), time of the glycan binding to the chromatography paper (X9), duration of acetonitrile wash (X10) and the storage time prior to the HPLC analysis (X11). High and low levels of all screened factors are shown in Table 6.

Table 6: Low and high values of all 11 factors, applied in 12 randomly conducted runs. The conditions of standard method procedure are shown in the last column.

Factors			Levels		
Mark	Description	**Measure**	**Low**	**High**	**SOP**
X1	DTT reduction	min	15	30	15
X2	Alkylation	min	30	90	30
X3	Gel freezing	/	YES	NO	YES
X4	Enzyme concentration	mU/µL	1,25	2,5	2,5
X5	Enzyme time exposure	h	16	24	16
X6	Second day storage	h	0	24	0
X7	Formic acid treatment	min	40	90	40
X8	2AB labeling duration	h	2	3,5	120
X9	Drying	min	5	15	15
X10	ACN wash	min	15	25	15
X11	Time before injection	h	0	48	48

X1: duration of the exposure of proteins to the reducing agent dithiotreitol; X2: duration of the exposure to the alkylating agent iodoacetamide; X3: whether the polyacrylamide gel blocks comprising proteins should be freezed; X4: concentration of N-glycosidase-F; X5: duration of the enzyme digestion; X6: storage of the eluted and dried glycans after the enzymatic digestion; X7: duration of the formic acid treatment; X8: duration of the 2-AB binding reaction; X9: time of the glycan binding to the chromatography paper; X10: duration of the acetonitrile wash; X11: the storage time prior to the HPLC analysis.

Runs given were conducted in a duplicate at the random order to nullify the effect of extraneous or nuisance variables. Average value obtained from these two values was used for further analysis of deviation from expected results.

3.9 MAIN EFFECT CALCULATION AND STATISTICAL ANALYSIS:

Average value for each peak area percent was calculated from two independently replicated experiments for each of the 12 runs (Table 5). For each experiment the same standardized plasma sample was used. The deviation from previously determined expected result was calculated (coefficient of variance) and used as response variable. Expected result for each peak area percent was obtained by analyzing standardized plasma sample following the laboratory standard operating procedure (SOP). In the course of this work, standardized plasma sample was analyzed 7 times according to SOP and the average value was used as expected peak area percentage.

Effects of the responses were calculated according to the equation (1).

$$E_x = \frac{2\sum Y_i^+}{n} - \frac{2\sum Y_i^-}{n} \quad (1)$$

where E represents effect of factor 1 to 7, designated with index x, n is the total number of experimental runs (12), Y_i^+ are the recorded coefficients of variance (CV) at upper level obtained for each factor, Y_i^- are the recorded CV at lower level obtained for the same factor.

The significance of calculated effects was first analyzed graphically by plotting a half normal probability plot. Effects falling out of the line represent the potentially significant effects.

To confirm significance of the effects, standard error was calculated from four smallest effects which were considered negligible or not important. Those effects were pooled for standard error calculation according to formula (2).

$$SE_e = \sqrt{\frac{\sum(E_x^n)^2}{N_n}} \quad (2)$$

where $\sum(E_x^n)^2$ is sum of squares of the N_n negligible effects. (3)

Absolute effect values were compared to E_{crit} values which were calculated according to the equations:

$$t_{crit} = TINV(\alpha, DF) \quad (4)$$

$$E_{crit} = t_{crit}\,(SE)_e \quad (5)$$

Where α indicates a level of significance (0.01) and DF indicates degrees of freedom in standard error estimation. An effect was considered significant at the α = 0.01 if:

$$|E_x| > E_{crit}$$

3.10 GENOME-WIDE ASSOCIATION ANALYSIS

Each trait was tested for normality within each cohort and, then the transformation that performed best for all cohorts was used. Models including sex, age and fibrinogen as covariates were tested for each cohort separately. Any covariate that was significant within any cohort was included as a covariate in the final model. Genome-wide associations were performed for all glycan measures using the same transformation to normality and covariates for each cohort separately then combined in a meta-analysis. The ‘‘mmscore’’ function of the GenABEL package for R statistical software[68] was used for the association test under an additive model. This score test for family based association takes into account pedigree structure and allowed unbiased estimations of SNP allelic effect when relatedness is present between examinees[69]. The relationship matrix used in this analysis was generated by the ‘‘ibs’’ function of GenABEL which used IBS genotype sharing to determine the realized pair wise kinship coefficient. Meta-analysis was performed using the MetABEL package for R[68]. An association was considered statistically significant at the genome-wide level if the p-value for an individual SNP was less than 5×10^{-8} (based on Bonferroni correction to account for multiple testing). All identified SNPs that reached significance or seemed to be suggestive of significance were visualized using Haploview software[70].

3.11 STATISTICAL ANALYSIS

Mann-Whitney U test, non-parametric statistical hypothesis test, was used for the comparison of all glycan traits. Significant level for p was corrected according to Bonferroni correction. The study was performed using 108 samples for GCK-MODY, 110 samples for HNF1A-MODY, 17 samples for HNF4A-MODY, 70 samples for Type 1 diabetes and 66 samples for Type 2 diabetes. Glycan indexes were tested for known significant covariates (sex, age).

Receiver operating characteristic curve (ROC) is used to estimate a new diagnostic test discriminatory power. Observed variable values should be on continuous scale and the plot is obtained by plotting specificity (true positive) versus sensitivity (true negative) of many randomly tested threshold points. Area under curve (AUC) was used as the measure of discriminatory power of promising glycan traits, displaying probability that when the variable is observed for a randomly selected subject from the diseased population and a randomly selected subject from the normal population the resulting values will be in the correct order. Thus, if subjects from the diseased population tend to have high values on the test, the ROC AUC is the probability that a randomly selected subject from the diseased population will have a higher value than a randomly selected subject from the normal population. If a test can discriminate perfectly, it will have an AUC of 1.0 while a test with no diagnostic capability has an AUC of 0.5[71].

4 RESULTS

4.1 ROBUSTNESS OF N-GLYCAN ANALYSIS

Plasma sample obtained after pooling the samples from 90 healthy individuals was used for the purpose of method robustness testing. In that way we obtained a standard sample which was used in all 12 runs of the experiment. Blood was drawn on anticoagulant, centrifuged at the time of collection and plasmas were stored at -20°C until analysis. The study was approved by the appropriate Ethical Committees in Croatia. N-glycans from the plasma proteins were released, fluorescently labeled and analyzed using hydrophilic interaction chromatography. Obtained chromatograms were separated into 19 peaks, each of them encompassing several structurally and compositionally similar oligosaccharides (last four peaks were summed together for the purpose of analysis, Figure 11). Value for each peak is presented as the area percentage of the total glycome of the analyzed sample.

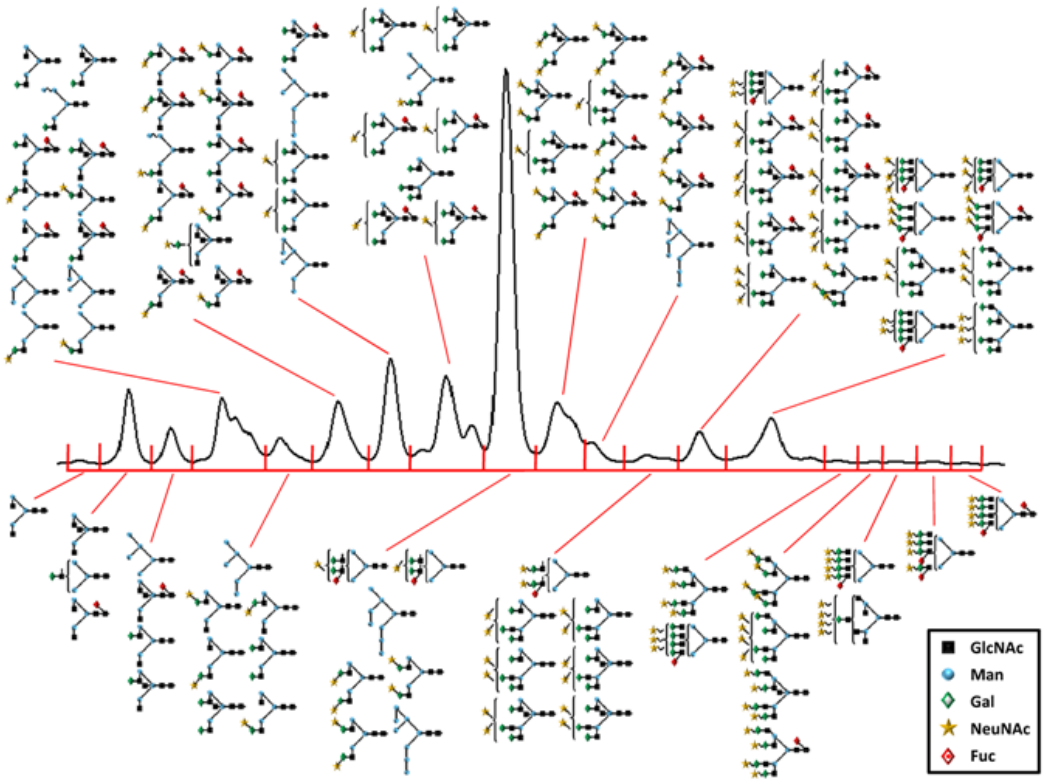


Figure 11: Schema of a typical plasma glycome structures separated into 16 chromatographic peaks by hydrophilic interaction chromatography (last four peaks were pooled together).

Structures represented by each peak were analyzed on the basis of their elution positions and measured in glucose units then compared to reference values in the

"Glyco- Base" database for structure assignment[65]. GU units were assigned according to 2AB labeled standard of glucose oligomers of different sizes. Typical chromatograms of initial (referred to as GP1–GP16) plasma N-glycome are shown in Figure 11. All analyzed peaks represent 16 different responses for eleven screened factors (Table 7). Each of the selected 12 runs (Table 5) was run in duplicate and obtained mean value was used for the calculation of variance which represents the aberration from the result obtained by standard operating procedure. After the half-normal probability plot analysis was done it seemed that some of the factors tested had significant influence on thirteen glycan groups (GP1, GP3, GP5, GP6, GP7, GP8, GP9, GP10, GP12, GP13, GP14, GP15 and GP16). Comparison of each factor absolute effect value on all 16 glycan groups with critical effect value (E_{crit}) calculated from the four lowest responses coincided with this finding and identified ten factors as having significant effects on thirteen glycan groups. Further on, calculated E_{crit} values (coefficient of variance) for all glycan groups reporting significant effects were compared with the coefficient of variance reported previously for the method[9].

Table 7: Factors of the high throughput method screened for the method robustness. Each factor is denoted wit X followed by the ordinal

Description	Factors
X1	DTT reduction
X2	Alkylation
X3	Gel freezing
X4	Enzyme
X5	Enzyme time
X6	Second day storage
X7	Formic acid
X8	2AB labeling
X9	Drying
X10	ACN wash
X11	Time before

All values smaller than previously reported[9] were regarded as negligible what resulted in removal of seven glycan groups from the list of the ones showing significant effects (Table 8). The half-normal probability plots of the glycan groups on which investigated

Table 8: Calculated absolute effect values of 11 factors on 16 responses (GP1 – GP16). E_{crit} – Effect critical value; GP1-GP16 – 16 separated glycan groups; CV_p - Coefficient of variance previously reported [15]

		X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	E_{crit}	CV_p
Absolute Effect values	GP1	2.16	1.17	0.21	1.2	5.54	0.35	1.46	3.83	1.2	2.02	1.2	2.4	7.39
	GP2	5.86	4.47	0.49	2.16	3.47	5.6	7.27	2.93	2.72	3.27	2.82	10.4	3.38
	GP3	2.58	1.17	0.62	0.5	1.04	0.52	0.25	0.08	0.18	0.25	1.37	0.95	2.6
	GP4	5.58	3.17	0.44	1.42	2.13	4.72	5.95	2.33	2.44	2.71	2.45	8.04	2.75
	GP5	0.85	1.91	0.43	0.59	0.61	0.96	0.39	0.41	0.21	0.65	1.86	1.71	3.98
	GP6	4.36	1.77	0.85	0.3	0.57	3.29	3.24	1.52	1.03	1.72	2.25	3.42	1.82
	GP7	0.67	1.21	0.18	0.73	0.17	0.37	0.5	0.65	0.09	0.41	0.42	1.04	10.1
	GP8	0.07	0.29	0.99	1.87	0.21	0.5	0.36	0.01	0.53	0.36	0.84	0.84	1.26
	GP9	0.35	0.9	0.04	0.65	0.84	1.31	1.09	0.64	0.16	0.73	0.07	0.91	1.37
	GP10	1.45	0.19	0.9	0.4	0.98	0.08	3.26	0.75	0.57	0.29	0.78	1.22	2.63
	GP11	1.31	1.04	0.4	0.99	3.26	1.93	1.96	2.33	1.14	4.0	3.18	4.32	3.33
	GP12	0.83	0.11	0.32	0.94	0.34	1.08	0.63	0.32	0.35	0.39	1.61	1.32	3.14
	GP13	2.99	1.17	1.91	0.56	1.06	0.2	1.55	2.65	2.13	1.71	0.28	2.87	2.85
	GP14	4.06	0.63	0.83	0.0	0.86	0.66	2.28	3.53	2.09	2.14	0.08	2.12	2.97
	GP15	2.56	1.0	0.46	2.04	7.93	2.39	3.33	7.94	2.95	9.73	0.45	5.44	6.64
	GP16	2.07	0.84	0.34	2.03	6.08	2.58	2.62	7.76	3.83	9.72	1.53	6.21	12.3

factors showed significant effects are shown in Figure 12. In total, eight factors (X1, X3, X4, X5, X7, X8, X10, X11) displayed significant effects on 6 (GP6, GP8, GP10, GP13, GP14, GP15) out of 16 responses (glycan groups) observed.

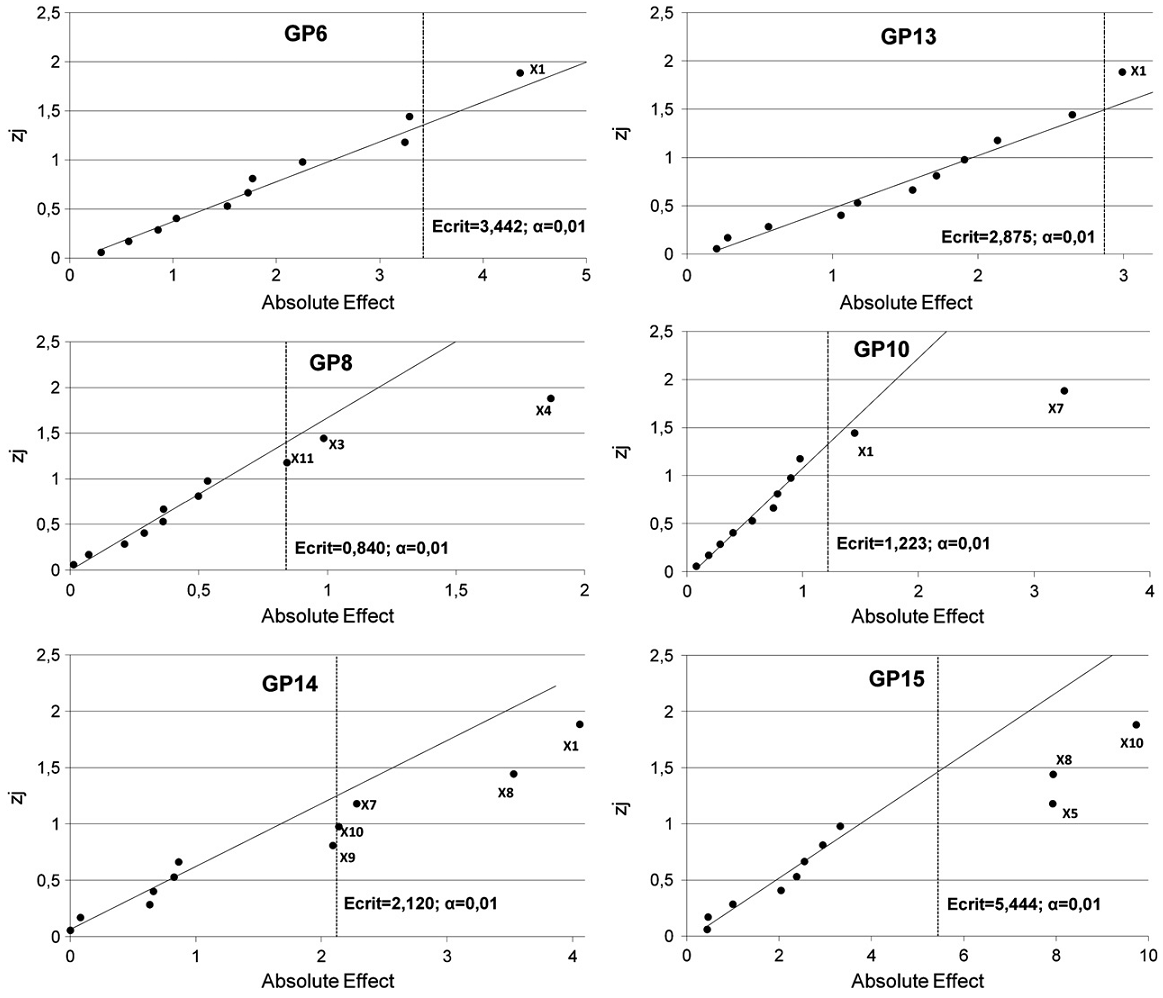


Figure 12: Half-normal probability plot for 11 factors for 6 responses (GP6, GP8, GP10; GP13, GP14, GP15) investigated in 12 runs by Plackett-Burman design of experiments. Ecrit values for 0.01 level of significance are marked by a dashed line. GP: group of glycans; X1: duration of the exposure of proteins to the reducing agent; X2: duration of the exposure to the alkylating agent iodoacetamide; X3: whether the polyacrylamide gel blocks comprising proteins should be freezed; X4: concentration of N-glycosidase-F; X5: duration of the enzyme digestion; X6: storage of the eluted and dried glycans after the enzymatic digestion; X7: duration of the formic acid treatment; X8: duration of the 2-AB binding reaction; X9: time of the glycan binding to the chromatography paper; X10: duration of acetonitrile wash; X11: the storage time prior to the HPLC analysis.

Factors X1 (time of exposure to DTT reduction) appears in four response variables (peaks) as significant. X1 displays significant effect for GP6, GP10, GP13 and GP14. Factor X3, gel freezing, appears significant for GP8. Factor X4 (enzyme concentration) appeared significant in the case of GP8. Although this factor showed statistical significance, the absolute values of the effect are slightly higher than previously reported and quite lower than overall variance reported[9].

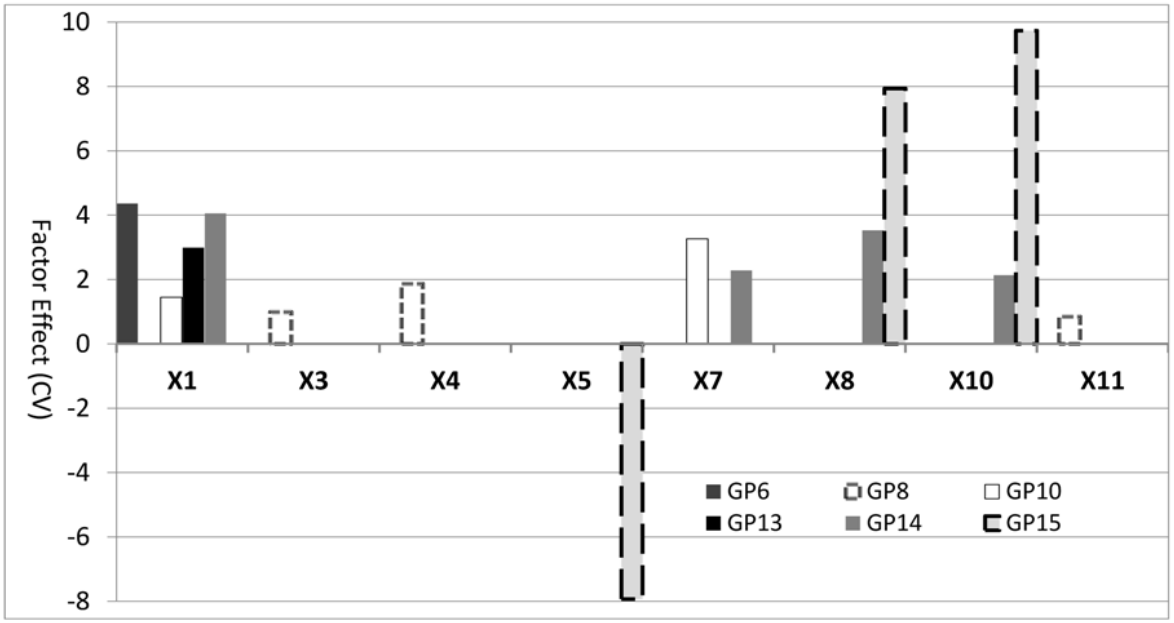


Figure 13: Effect values for 8 significant factors represented by vertical columns. X1: duration of the exposure of proteins to the reducing agent; X3: whether the polyacrylamide gel blocks comprising proteins should be freezed; X4: concentration of N-glycosidase-F; X5: duration of the enzyme digestion; X7: duration of the formic acid treatment; X8: duration of the 2-AB binding reaction; X10: duration of acetonitrile wash; X11: the storage time prior to the HPLC analysis.

Factor X5, time of exposure to enzyme, emerged significant for response GP 15 and factor X7, formic acid treatment, for GP10 and GP14. Factor X8, duration of 2AB labeling, and X10, duration of acetonitrile wash, appear significant in the case of two response variables (GP14 and GP15). Factor X11, time of storage prior to HPLC analysis, displayed significance for the GP8 (Figure 13).

Time of exposure to DTT has the highest number of response variables which showed significant effect on variability of glycan analysis and should be therefore seen as the most critical step in the high throughput method and should be carried out with extreme precaution. Considering low and high levels of factors tested (15 minutes vs. 25 minutes) and effect direction, it is evident that the reduction of proteins should not be prolonged more than 15 minutes otherwise higher than usual variability will be recorded.

Glycan peaks GP8 and GP15 and glycan peak GP14 showed significant effects for three and four factors, respectively. They represent the most sensitive groups of glycans regarding method conditions tested in this work.

4.2 N-GLYCAN ANALYSIS OF THREE POPULATION COHORTS

Method for the release, labeling and analysis of plasma glycans was optimized according to the findings from robustness experiments. Glycoms of in total 2559 samples from three different population cohorts, CROATIA-VIS, CROATIA-KORCULA and ORCADES, were analyzed. N-glycans were released from plasma proteins by the digestion with N-glycosidase F, labeled with 2-aminobenzamide (2-AB) and finally analyzed by hydrophilic interaction high performance liquid chromatography (16 glycan peaks denoted as GP1-GP16). Additionally, released and 2-AB labeled glycans were treated with *A. Ureafaciens* sialidase in order to release α2-3, α2-6 and α2-8 linked sialic acid and subjected to second hydrophilic chromatography analysis (13 glycan peaks denoted as DG1-DG13). Glycan structures covered by 29 peaks are listed in Table 3 and Table 4. Values of individual peaks are presented as the relative percentage area of the total glycome and for both chromatography analyses the percentages add to 100%. To obtain clearer insight about general glycan features (fucosylation, sialylation ect.) another additional glycan parameters were introduced which were calculated from measured data (Table 12-14). Descriptive statistic for all three populations, including minimum, maximum, 1st quartile, 3th quartile and median is shown in Tables 9-11.

Table 9: Minimum, maximum, median and interquartile range for all measured glycan traits from the CROATIA-KORCULA cohort.

	KORCULA				
	min	**1st quartile**	**median**	**3th quartile**	**max**
GP1	0.02	0.09	0.13	0.18	0.92
GP2	1.06	2.56	3.27	4.09	11.09
GP3	0.84	1.73	1.98	2.24	5.02
GP4	1.99	4.26	4.94	5.71	11.07
GP5	1.55	2.19	2.41	2.71	4.96
GP6	1.46	3.01	3.48	4.07	6.75
GP7	6.34	9.45	10.41	11.56	29.18
GP8	5.46	8.06	8.94	9.95	15.28
GP9	27.84	36.13	38.13	40.18	53.77
GP10	3.34	6.54	7.46	8.40	14.84
GP11	0.63	1.36	1.60	1.90	3.45
GP12	0.65	1.55	1.79	2.27	3.72
GP13	0.98	4.53	5.50	6.48	11.18
GP14	2.94	5.54	6.53	7.58	13.44
GP15	0.07	0.47	0.60	0.73	1.25
GP16	0.27	0.89	1.06	1.23	2.53
DG1	0.04	0.13	0.17	0.22	0.90
DG2	1.26	2.80	3.47	4.25	10.97
DG3	1.50	2.39	2.68	2.99	5.44
DG4	2.33	5.47	6.16	6.93	13.75
DG5	36.27	47.97	50.30	52.77	69.85
DG6	6.95	12.36	14.00	15.82	23.90
DG7	0.70	1.30	1.54	1.83	5.68
DG8	5.07	10.45	11.87	13.36	19.42
DG9	0.30	2.81	3.66	4.67	9.65
DG10	0.50	1.04	1.16	1.27	2.25
DG11	0.94	2.08	2.44	2.79	4.38
DG12	0.10	0.57	0.75	0.96	2.42
DG13	0.20	0.61	0.75	0.91	2.22

Table 10: Minimum, maximum, median and interquartile range for all measured glycan traits from the CROATIA-VIS cohort

	VIS				
	min	**1st quartile**	**median**	**3th quartile**	**max**
GP1	0.03	0.11	0.16	0.22	0.83
GP2	0.45	3.06	3.93	5.01	17.90
GP3	0.72	1.81	2.10	2.39	5.97
GP4	0.62	5.17	5.89	6.70	11.21
GP5	1.32	2.01	2.24	2.51	5.38
GP6	0.87	3.52	4.03	4.61	7.94
GP7	6.16	9.12	10.04	11.32	30.30
GP8	5.75	8.96	9.83	10.95	17.47
GP9	24.29	35.07	36.81	38.83	48.53
GP10	3.50	6.35	7.35	8.52	18.75
GP11	0.94	1.90	2.22	2.64	5.70
GP12	0.70	1.30	1.46	1.65	4.18
GP13	0.59	3.93	4.86	5.73	9.66
GP14	2.51	5.35	6.23	7.22	13.05
GP15	0.06	0.33	0.44	0.55	1.51
GP16	0.11	0.51	0.64	0.78	2.15
DG1	0.04	0.12	0.17	0.23	0.82
DG2	0.56	3.14	3.99	5.06	18.00
DG3	1.04	2.38	2.68	3.03	6.68
DG4	0.99	6.14	6.98	7.82	12.37
DG5	33.73	47.07	49.14	51.22	64.47
DG6	6.24	14.08	15.71	17.74	34.41
DG7	0.69	1.26	1.50	1.77	3.79
DG8	4.80	9.50	10.92	12.20	17.26
DG9	0.17	2.72	3.60	4.55	8.68
DG10	0.63	0.96	1.07	1.18	1.85
DG11	0.79	1.81	2.13	2.49	4.95
DG12	0.04	0.51	0.68	0.88	2.46
DG13	0.18	0.40	0.50	0.68	2.80

Table 11: Minimum, maximum, median and interquartile range for all measured glycan traits from the ORCADES cohort

	ORCADES				
	min	**1st quartile**	**median**	**3th quartile**	**max**
GP1	0.02	0.11	0.15	0.22	0.89
GP2	1.39	3.13	4.29	5.62	10.27
GP3	0.78	1.76	2.15	2.61	5.35
GP4	3.11	5.81	6.94	8.29	14.64
GP5	1.06	2.38	3.12	4.53	11.82
GP6	2.11	4.23	5.01	6.10	9.95
GP7	6.90	9.75	11.75	14.73	20.66
GP8	7.18	9.98	10.96	12.10	16.57
GP9	20.66	31.19	34.90	37.62	46.18
GP10	2.74	5.60	6.46	7.42	11.03
GP11	0.82	2.04	2.31	2.56	3.60
GP12	0.27	1.33	1.69	2.14	5.18
GP13	0.34	2.36	3.44	4.96	11.37
GP14	0.34	2.08	3.25	5.28	12.22
GP15	0.00	0.14	0.27	0.50	1.78
GP16	0.03	0.18	0.35	0.82	3.70
DG1	0.05	0.12	0.15	0.19	0.74
DG2	1.46	2.82	3.65	4.43	13.84
DG3	1.06	2.01	2.29	2.57	5.49
DG4	3.31	6.14	6.99	7.79	12.65
DG5	37.46	46.94	48.72	50.50	58.70
DG6	8.64	13.94	15.50	17.35	24.82
DG7	0.65	1.12	1.29	1.51	2.94
DG8	5.26	11.10	12.31	13.63	20.98
DG9	0.79	2.52	3.39	4.44	10.21
DG10	0.64	0.94	1.05	1.17	1.94
DG11	0.86	2.02	2.49	3.11	6.11
DG12	0.10	0.47	0.68	0.94	2.60
DG13	0.17	0.50	0.69	0.93	3.54

Table 12: Calculated glycan features for CROATIA-KORCULA cohort: Core fucosylated glycans (FUC-C); Antennary fucosylated glycans (FUC-A); Biantennary glycans (BA); Monosialylated biantennary glycans (BAMS); Disialylated biantennary glycans (BADS); Triantennary glycans (TRIA); Nongalactosylated glycans (G0); Monogalactosylated glycans (G1); Digalactosylated glycans (G2); Trigalactosylated glycans (G3); Tetragalactosylated glycans (G4).

	KORČULA				
	min	**1st quartile**	**median**	**3th quartile**	**max**
BAMS	19.95	28.22	29.87	31.96	47.31
BADS	54.04	69.11	71.78	74.08	84.43
BA	70.98	77.34	78.98	80.87	89.68
TRIA	8.79	15.53	16.98	18.21	24.34
TA	1.54	3.53	3.99	4.52	7.17
C-FUC	9.60	19.19	21.59	24.73	35.40
A-FUC	1.34	2.49	2.97	3.51	11.05
A2	0.03	0.12	0.15	0.20	0.91
G0	1.39	2.98	3.64	4.45	11.65
G1	5.00	8.06	8.91	9.78	19.19
G2	56.43	64.49	66.25	67.97	82.11
G3	6.24	12.66	14.26	15.61	22.54
G4	0.43	1.40	1.66	1.95	3.29

Table 13: Calculated glycan features for CROATIA-VIS cohort: Core fucosylated glycans (FUC-C); Antennary fucosylated glycans (FUC-A); Biantennary glycans (BA); Monosialylated biantennary glycans (BAMS); Disialylated biantennary glycans (BADS); Triantennary glycans (TRIA); Nongalactosylated glycans (G0); Monogalactosylated glycans (G1); Digalactosylated glycans (G2); Trigalactosylated glycans (G3); Tetragalactosylated glycans (G4).

	VIS				
	min	**1st quartile**	**median**	**3th quartile**	**max**
BAMS	22.51	28.66	30.51	32.49	51.67
BADS	50.04	67.82	70.49	72.92	84.72
BA	68.08	79.39	80.80	82.42	88.79
TRIA	9.14	14.51	15.64	16.82	24.15
TA	1.75	2.92	3.36	3.88	9.21
C-FUC	9.80	21.65	24.35	26.92	50.50
A-FUC	1.12	2.51	2.94	3.52	7.41
A2	0.04	0.12	0.17	0.22	0.83
G0	0.60	3.31	4.15	5.24	18.19
G1	2.03	8.77	9.65	10.59	17.01
G2	56.62	64.93	66.68	68.65	76.60
G3	6.23	11.55	12.61	13.91	21.82
G4	0.24	0.88	1.09	1.31	3.25

Table 14: Calculated glycan features for ORCADES cohort: Core fucosylated glycans (FUC-C); Antennary fucosylated glycans (FUC-A); Biantennary glycans (BA); Monosialylated biantennary glycans (BAMS); Disialylated biantennary glycans (BADS); Triantennary glycans (TRIA); Nongalactosylated glycans (G0); Monogalactosylated glycans (G1); Digalactosylated glycans (G2); Trigalactosylated glycans (G3); Tetragalactosylated glycans (G4).

	ORCADES				
	min	**1st quartile**	**median**	**3th quartile**	**max**
BAMS	23.49	31.48	35.00	39.90	49.01
BADS	43.12	60.47	67.09	71.43	84.56
BA	66.37	77.17	79.14	80.92	87.36
TRIA	10.90	15.66	16.80	18.06	23.28
TA	1.49	3.17	3.97	4.90	10.59
C-FUC	13.00	21.87	24.18	26.60	37.45
A-FUC	1.26	2.24	2.60	3.03	5.42
A2	0.04	0.12	0.16	0.21	0.78
G0	1.55	2.98	3.82	4.61	14.09
G1	5.13	8.31	9.31	10.20	15.36
G2	55.70	63.95	65.71	67.38	74.82
G3	1.37	6.33	9.00	12.46	21.89
G4	0.06	0.32	0.63	1.35	5.48

4.3 N-GLYCOME GENOME-WIDE ASSOCIATION STUDY

Genome wide association study (GWAS) was performed at Medical Research Council Human Genetics Unit of Institute of Genetics and Molecular Medicine in Western General Hospital, Edinburgh in a scope of a scientific collaboration. A meta-analysis of GWAS data was conducted for collected plasma N-glycan traits measured in three population-based cohorts, CROATIA-VIS (n = 960), CROATIA-KORCULA (n = 967) and ORCADES (n = 502). Additive SNP effects were tested in each cohort independently and then combined in an inverse-variance weighted meta-analysis. The genome-wide significance threshold for the meta-analysis was set at $5x10^{-8}$.

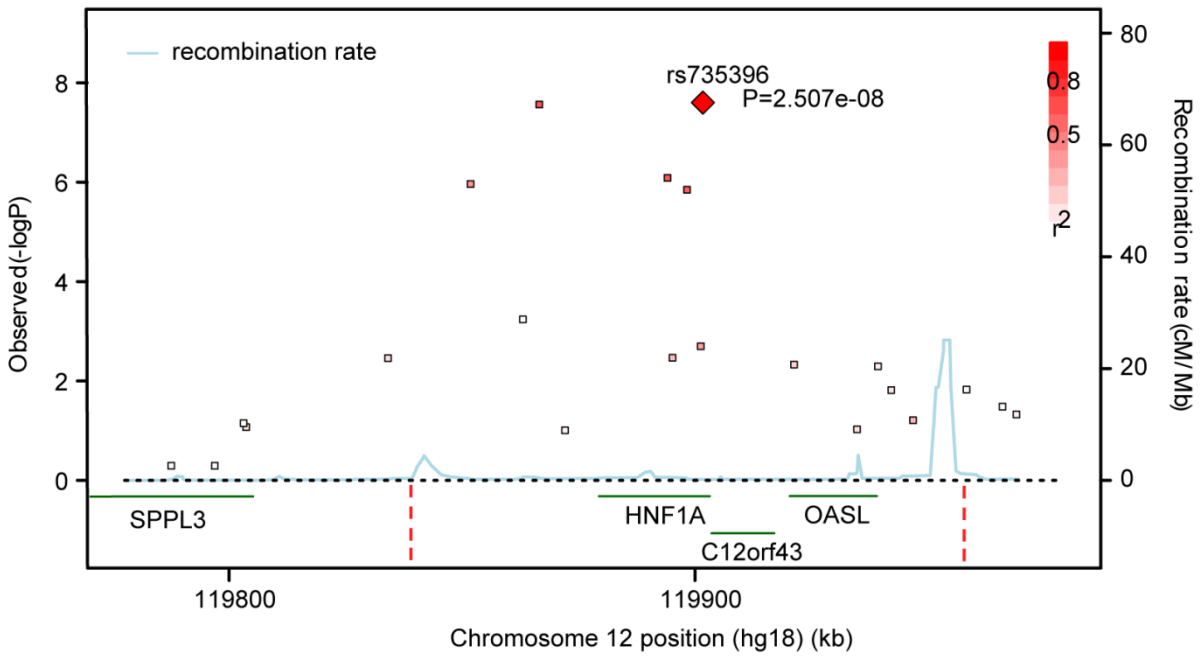


Figure 14: Significance plot for the region of chromosome 12 from the meta-analysis. Blue line represents recombination rate with the scale on the right axes. Red diamond indicates the SNP with the most significant association with the HNF1A region with indicated P value. Different shades of red represent the degree of linkage disequilibrium between the most significant SNP (red diamond) and all tested SNPs[27].

Genome-wide significant associations were found for DG1, DG6, DG7, DG9, DG11, DG12 as well as FUC-A (Table 15, page 24). Two SNPs on chromosome 12, rs7953249 and rs735396, showed genome-wide significant associations with DG7 ($p = 1.97x10^{-8}$, $p = 1.75x10^{-8}$). The latter SNP was also associated with DG11 ($p = 4.44x10^{-8}$), with an effect in the opposite direction. Both SNPs are located in the HNF1A (Entrez GeneID: 6927) gene region: rs7953249 is found 13 kb 59 to the gene and rs735396 is in intron 9. Association profile for DG7, representing antennary fucosylated glycan structures, is represented in its genomic context in Figure 14 for the associated region. Quantile–quantile plot for DG7 association were consistent with an excess of true genetic associations, with modest genomic control inflation[72–74] for each population (inflation factor, 1.04 for each population and the meta-analysis), suggesting that the observed results were not due to population stratification[72] (Figure 15).

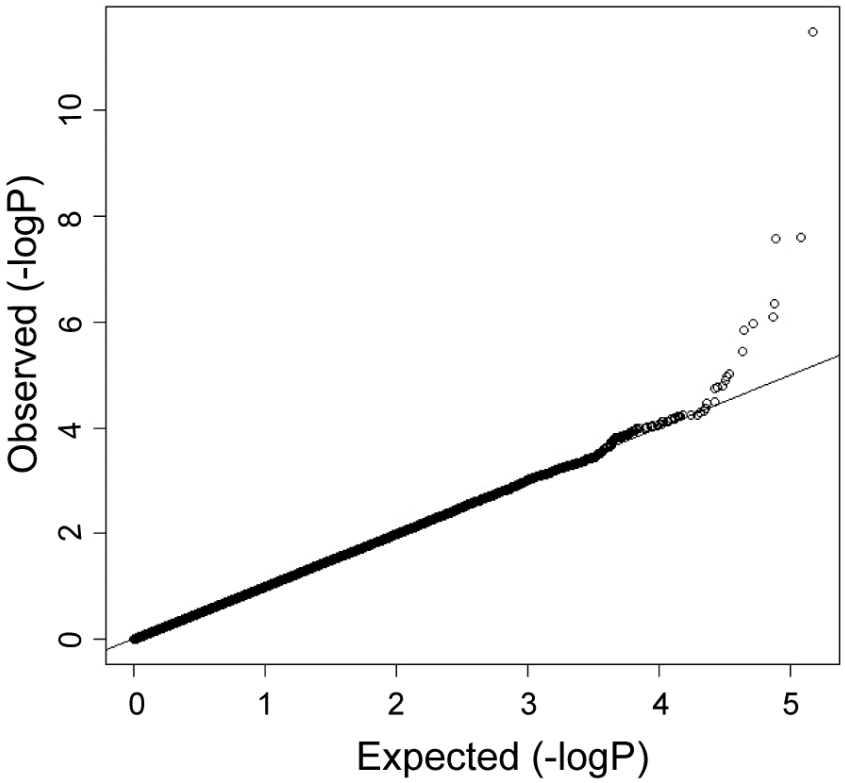
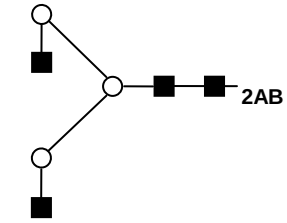
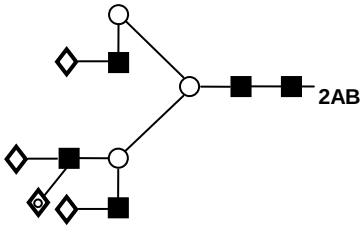
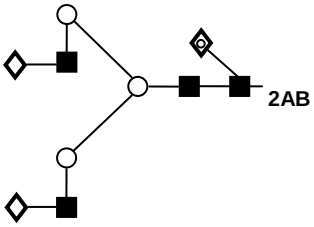
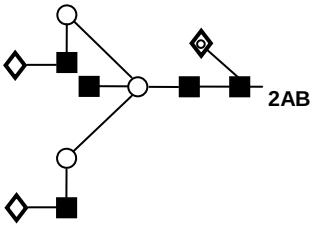
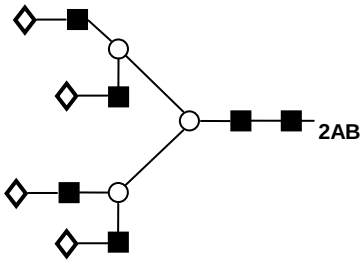
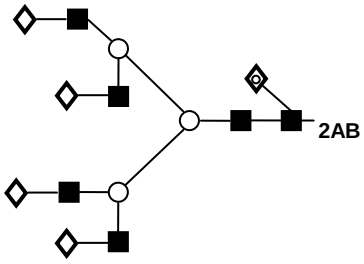
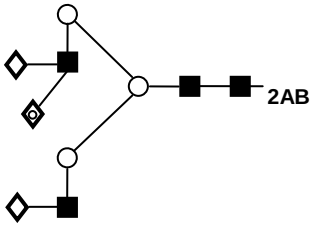
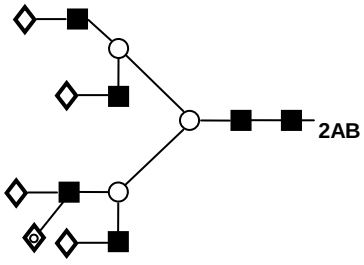


Figure 15: Quantile-quantile plot for the test statistic of the most significant SNP (rs735396) associated with DG7 peak, dominantly glycan structures with antennary fucosylation. $-Log_{10}$ of the association p-values are plotted against $-log_{10}$ of the expected p-value under the null hypothesis of no association for the meta-analysis[27].

The effect size of the G allele of rs735396 is - 0.1767 (standard deviation units, after adjustment for sex, age and fibrinogen; s.e. 0.0314) for DG7, which only contains glycans with antennary fucose, and in the opposite direction (0.1699 standard deviation units, after adjust-ment for age and fibrinogen; s.e. 0.0310) for DG11, which has no antennary fucose (Table 15). All significant SNPs in this region had a similar effect size (absolute value of the range: 0.1396–0.1767), representing 1–3% of the trait variance. Comparison of models including rs7953249 and rs735396 separately and combined suggests that the causal variant is located between these two SNPs. This was confirmed by analysis of imputed data based on HapMap release 2 with the most significant SNPs located across intron 1 of HNF1a.

Table 15: Glycan structures are presented schematically to denote the construction of the structure. ■ - N-acetylglucosamine, ○ - mannose, ◇ - galactose, ◈ - fucose; 2AB – 2-aminobenzamide

Peak	Main complex glycan structures present	Peak	Main complex glycan structures present
DG1	 A2	DG9	 A3F1G3
DG6	 FA2G2  FA2BG2	DG11	 A4G4  FA4G4
DG7	 A2F1G2	DG12	 A4F1G4

4.4 ANALYSIS OF DIABETES COHORTS

4.4.1 HNF1A-MODY N-glycan analysis.

In total 346 samples of several diabetes subtypes were analyzed: 110 patients with the diagnosed HNF1A MODY subtype [n(serum) = 49, n(plasma) = 61], 17 patients with the HNF4A MODY diagnosis [n(serum) = 31, n(plasma) = 77], 108 patients with the GCK MODY diagnosis [n(serum) = 10, n(plasma) = 7] 69 patients with type 1 diagnosis [n(serum) = 0, n(plasma) = 69], and 91 patients with type 2 diagnoses [n(serum) = 16, n(plasma) = 75]. N-glycans were released from plasma proteins by the digestion with N-glycosidase F, labeled with 2-aminobenzamide (2-AB) and finally analyzed by hydrophilic interaction high performance liquid chromatography (16 glycan peaks denoted as GP1-GP16). Additionally, released and 2-AB labeled glycans were treated with *A. Ureafaciens* sialidase in order to release α2-3, α2-6 and α2-8 linked sialic acid and subjected to second hydrophilic chromatography analysis (13 glycan peaks denoted as DG1-DG13). Glycan structures covered with 29 peaks are listed in table 3 and table 4. Values of individual peaks are presented as the relative percentage area of the total glycome and for both chromatography analyses the percentages add to 100%. Descriptive statistic for all three populations, including minimum, maximum, 1st quartile, 3th quartile and median is shown in tables 16-20.

Table 16: Minimum, maximum, median and interquartile range for all measured glycan traits from the diabetes type 2 cohort; GP1-GP16 – chromatography peaks of the initial glyco-profile. DG1-DG13 – chromatography peaks of the desialylated initial glycol profile. A-Fuc – antennary fucosylation.

	Type 2				
	min	**1st quartile**	**median**	**3th quartile**	**max**
GP1	0.04	0.07	0.11	0.16	0.50
GP2	1.17	2.35	3.26	4.22	7.50
GP3	1.11	1.54	1.87	2.18	2.96
GP4	2.02	3.57	4.40	5.45	8.64
GP5	1.52	1.78	1.96	2.16	3.00
GP6	1.99	2.89	3.40	3.90	6.07
GP7	6.08	8.12	8.83	9.89	12.20
GP8	7.56	9.12	9.73	10.64	13.41
GP9	31.75	37.68	39.82	41.74	47.40
GP10	4.55	6.43	7.22	8.18	10.11
GP11	1.17	1.82	2.14	2.35	3.36
GP12	1.42	1.92	2.14	2.47	4.05
GP13	2.70	4.75	5.79	6.59	13.07
GP14	4.00	6.06	7.27	8.15	12.20
GP15	0.06	0.35	0.43	0.57	1.16
GP16	0.34	0.70	0.86	1.09	1.80
DG1	0.05	0.10	0.15	0.20	0.37
DG2	1.26	2.41	3.31	4.23	7.27
DG3	1.24	1.92	2.33	2.68	3.56
DG4	3.33	4.84	5.87	6.95	9.97
DG5	42.32	48.44	50.70	52.20	60.28
DG6	8.21	11.28	12.73	15.08	21.07
DG7	0.78	1.68	2.14	2.79	4.71
DG8	7.30	11.17	12.53	14.57	23.41
DG9	0.48	2.83	4.04	5.11	7.60
DG10	0.82	1.08	1.23	1.42	2.31
DG11	1.28	2.04	2.42	2.78	4.62
DG12	0.11	0.53	0.68	0.84	1.48
DG13	0.41	0.67	0.82	1.02	1.87
A-FUC	1.55	3.20	4.09	5.30	8.73

Table 17: Minimum, maximum, median and interquartile range for all measured glycan traits from the diabetes type 1 cohort; GP1-GP16 – chromatography peaks of the initial glycol-profile. DG1-DG13 – chromatography peaks of the desialylated initial glycol- profile. A-Fuc – antennary fucosylation.

	Type 1				
	min	**1st quartile**	**median**	**3th quartile**	**max**
GP1	0.03	0.0725	0.1	0.1275	0.32
GP2	1.05	2.1025	2.61	3.0375	8.84
GP3	1.27	1.63	1.91	2.15	4.48
GP4	2.61	4.445	5.305	5.7575	9.6
GP5	1.56	2.0325	2.24	2.545	3.36
GP6	2.62	3.8425	4.33	4.8725	6.57
GP7	7.05	8.615	9.39	10.0675	12.77
GP8	7.98	10.185	11.18	12.125	14.41
GP9	30.22	35.0425	36.565	38.215	43.41
GP10	5.1	6.875	7.69	8.5375	10.92
GP11	1.15	1.4325	1.61	1.83	2.33
GP12	1.24	1.8225	2.015	2.2875	3.65
GP13	2.39	4.47	5.28	7.135	10.75
GP14	4.37	5.695	6.62	7.665	10.33
GP15	0.21	0.5225	0.635	0.7775	1.1
GP16	0.63	1.015	1.165	1.3475	1.92
DG1	0.04	0.07	0.11	0.14	0.31
DG2	1.18	2.3025	2.85	3.2875	8.74
DG3	1.65	2.18	2.41	2.735	5.23
DG4	3.85	5.73	6.575	7.16	10.94
DG5	43.4	47.5325	49.065	50.9925	55.26
DG6	10.48	13.5375	15.14	16.965	20.59
DG7	1.04	1.3225	1.65	1.985	2.69
DG8	6.87	10.0475	11.74	14.0175	20.16
DG9	1.05	3.1775	3.845	4.9425	7.89
DG10	0.75	0.95	1.065	1.14	1.53
DG11	1.46	2.175	2.475	2.9075	4.32
DG12	0.16	0.6125	0.8	1.185	1.84
DG13	0.45	0.6	0.76	1.025	2.54
A-FUC	1.954887	2.665489	3.327164	3.96092	4.878491

Table 18: Minimum, maximum, median and interquartile range for all measured glycan traits of the HNF4A-MODY cohort; GP1-GP16 – chromatography peaks of the initial glyco profile. DG1-DG13 – chromatography peaks of the desialylated initial glyco- profile. A-Fuc – antennary fucosylation.

	HNF4A-MODY				
	min	**1st quartile**	**median**	**3th quartile**	**max**
GP1	0.06	0.07	0.12	0.16	0.22
GP2	1.50	2.10	2.85	4.07	5.43
GP3	1.22	1.56	1.88	2.15	2.78
GP4	2.89	4.26	4.74	5.58	6.75
GP5	1.50	1.84	1.95	2.26	2.79
GP6	2.57	3.48	3.71	4.11	5.55
GP7	6.91	7.80	8.35	8.72	10.04
GP8	6.91	9.69	10.83	11.79	12.24
GP9	33.45	35.85	37.64	39.24	46.31
GP10	4.90	6.70	7.37	7.88	12.66
GP11	1.13	1.36	1.85	2.37	2.93
GP12	1.44	2.01	2.19	2.48	3.01
GP13	3.38	5.23	6.78	8.37	9.13
GP14	4.79	5.63	7.00	8.66	9.51
GP15	0.32	0.55	0.76	0.87	1.20
GP16	0.77	1.16	1.46	1.56	1.86
DG1	0.06	0.09	0.12	0.17	0.23
DG2	1.74	2.27	2.91	4.12	5.72
DG3	1.38	1.88	2.23	2.48	3.50
DG4	3.81	5.44	6.35	6.85	8.13
DG5	43.26	47.43	49.59	50.76	55.38
DG6	10.16	13.13	14.45	14.94	19.05
DG7	0.96	1.18	1.28	1.77	2.18
DG8	8.92	11.58	13.76	15.98	18.27
DG9	1.09	2.56	3.44	4.63	6.64
DG10	0.79	1.01	1.10	1.26	1.64
DG11	1.81	2.10	3.41	3.85	4.63
DG12	0.22	0.51	0.76	1.02	1.35
DG13	0.54	0.69	0.85	0.94	3.22
A-FUC	1.86	2.30	2.59	3.35	4.43

Table 19: Minimum, maximum, median and interquartile range for all measured glycan traits of the HNF1A-MODY cohort; GP1-GP16 – chromatography peaks of the initial glyco profile. DG1-DG13 – chromatography peaks of the desialylated initial glyco-profile; A-Fuc – antennary fucosylation.

	HNF1A-MODY				
	min	1st quartile	median	3th quartile	max
GP1	0.03	0.07	0.10	0.13	0.38
GP2	1.10	2.25	2.89	3.46	9.59
GP3	1.20	1.61	1.79	2.04	3.58
GP4	2.70	4.41	5.21	5.89	8.03
GP5	1.66	1.98	2.17	2.46	3.86
GP6	2.24	3.60	4.19	4.73	6.05
GP7	6.29	7.94	8.62	9.59	11.95
GP8	8.67	10.79	11.72	12.57	15.88
GP9	27.22	35.39	38.17	40.27	45.97
GP10	4.28	6.04	6.78	7.56	9.68
GP11	1.05	1.54	1.88	2.27	3.20
GP12	1.51	2.05	2.30	2.50	3.94
GP13	2.79	6.26	7.07	8.04	11.36
GP14	2.68	3.76	4.54	5.54	10.79
GP15	0.27	0.62	0.71	0.84	1.34
GP16	0.49	0.95	1.07	1.30	2.11
DG1	0.04	0.07	0.11	0.15	0.40
DG2	1.09	2.48	3.03	3.61	9.06
DG3	1.67	2.07	2.38	2.60	4.40
DG4	3.82	5.82	6.61	7.38	9.75
DG5	39.95	49.21	51.14	52.90	60.92
DG6	8.74	12.99	14.73	16.66	20.99
DG7	0.51	0.86	0.99	1.14	2.68
DG8	6.66	12.94	14.20	15.92	22.30
DG9	0.17	0.79	1.25	1.91	8.36
DG10	0.56	0.80	0.90	1.00	1.31
DG11	1.34	2.43	2.74	3.23	5.23
DG12	0.06	0.14	0.25	0.42	1.43
DG13	0.33	0.50	0.62	0.79	1.66
A-FUC	0.89	1.65	1.89	2.23	5.47

Table 20: Minimum, maximum, median and interquartile range for all measured glycan traits of the GCK-MODY cohort; GP1-GP16 – chromatography peaks of the initial glyco profile. DG1-DG13 – chromatography peaks of the desialylated initial glyco- profile. A-Fuc – antennary fucosylation.

	GCK-MODY				
	min	**1st quartile**	**median**	**3th quartile**	**max**
GP1	0.02	0.07	0.10	0.15	0.52
GP2	1.55	2.76	3.53	3.99	6.67
GP3	1.30	1.69	1.85	2.16	3.57
GP4	3.11	5.02	5.70	6.59	8.89
GP5	1.65	2.04	2.23	2.44	3.33
GP6	3.09	3.89	4.36	4.95	6.30
GP7	5.92	7.59	8.28	9.06	14.28
GP8	8.80	10.71	11.54	12.53	15.59
GP9	31.45	36.25	37.47	39.41	44.06
GP10	4.24	5.86	6.63	7.80	10.35
GP11	1.26	1.97	2.19	2.37	3.16
GP12	1.24	1.73	1.90	2.09	2.97
GP13	2.17	4.35	5.27	6.20	9.17
GP14	2.61	5.45	6.11	6.95	10.24
GP15	0.17	0.40	0.52	0.64	1.10
GP16	0.61	0.87	1.00	1.15	1.65
DG1	0.05	0.09	0.13	0.18	0.45
DG2	1.68	2.95	3.65	4.11	6.66
DG3	1.76	2.17	2.43	2.80	4.04
DG4	4.40	6.40	7.19	8.18	10.53
DG5	43.89	47.39	49.83	51.46	58.16
DG6	11.00	14.00	15.59	16.74	21.78
DG7	0.79	1.22	1.56	1.76	2.87
DG8	7.07	9.64	11.13	12.36	17.30
DG9	0.27	2.49	3.64	4.51	7.98
DG10	0.64	0.91	1.03	1.15	1.40
DG11	1.23	1.83	2.08	2.40	3.64
DG12	0.11	0.45	0.69	0.88	1.67
DG13	0.38	0.71	0.87	1.04	4.88
A-FUC	1.45	2.45	3.05	3.54	5.71

In the scope of this work both serum and plasma samples were analyzed. Potential differences in glycan structure composition between these two sample types were tested and are graphically represented in box plots (Figure 16-19). Mann-Whitney test was performed to define whether there is significant difference between glycan structures in two different sample types. P value of 0.05 for the statistically significant difference was corrected according to Bonferroni correction to p=0.0017. Only four peaks, GP1, GP3, GP7 and GP11, showed a statistically significant difference in the median values between two groups (Table 21). Three significant peaks, GP3, GP7 and GP11, contain fucosylated glycan structures and their values should be avoided when pooling together plasma and serum samples. Other 25 glycan values showed no statistically significant difference between serum and plasma samples and therefore can be pooled together and used for determining the difference between diabetes subtypes cohorts.

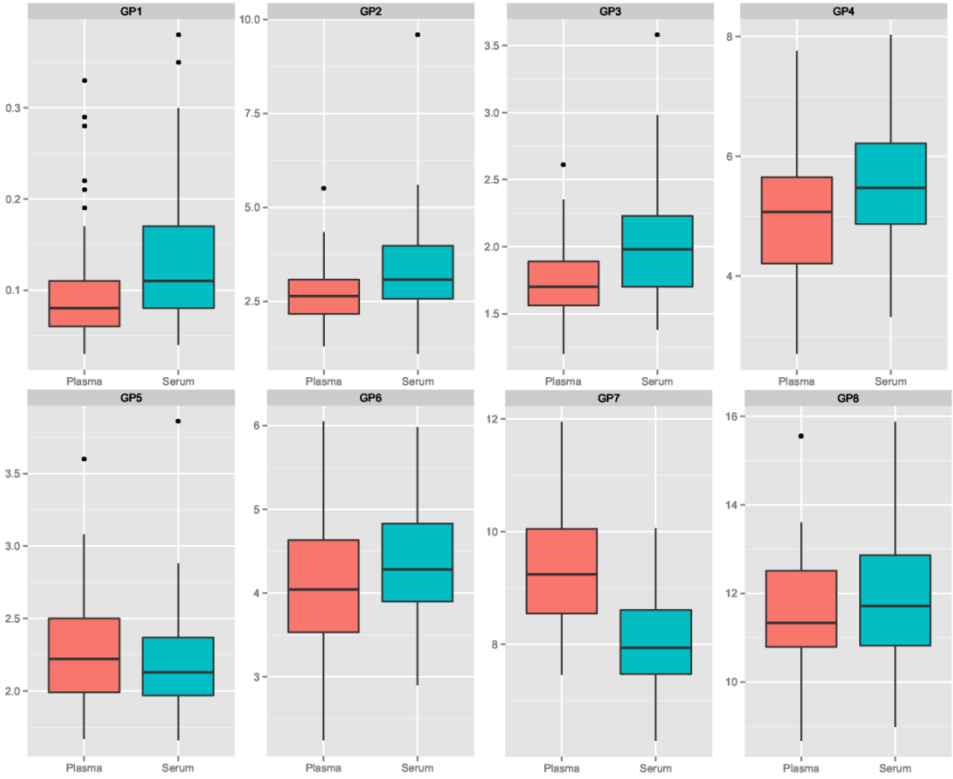


Figure 16: Box plots comparing plasma versus serum glycan values in the HNF1A-MODY cohort. Values for the peaks GP1-GP8.

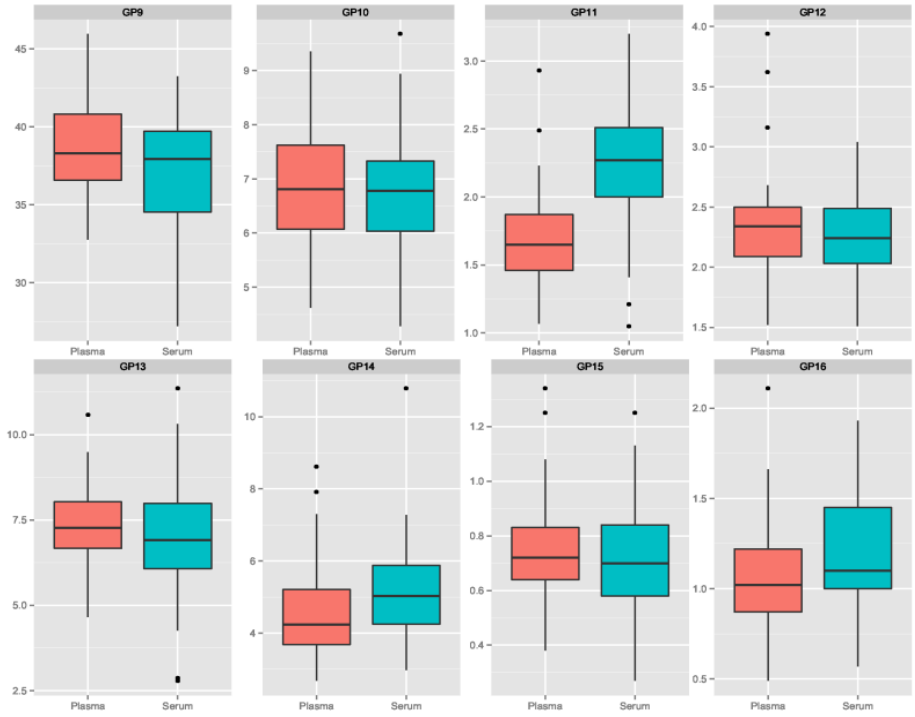


Figure 17: Box plots comparing plasma versus serum glycan values in the HNF1A-MODY cohort. Values for the peaks GP9-GP16.

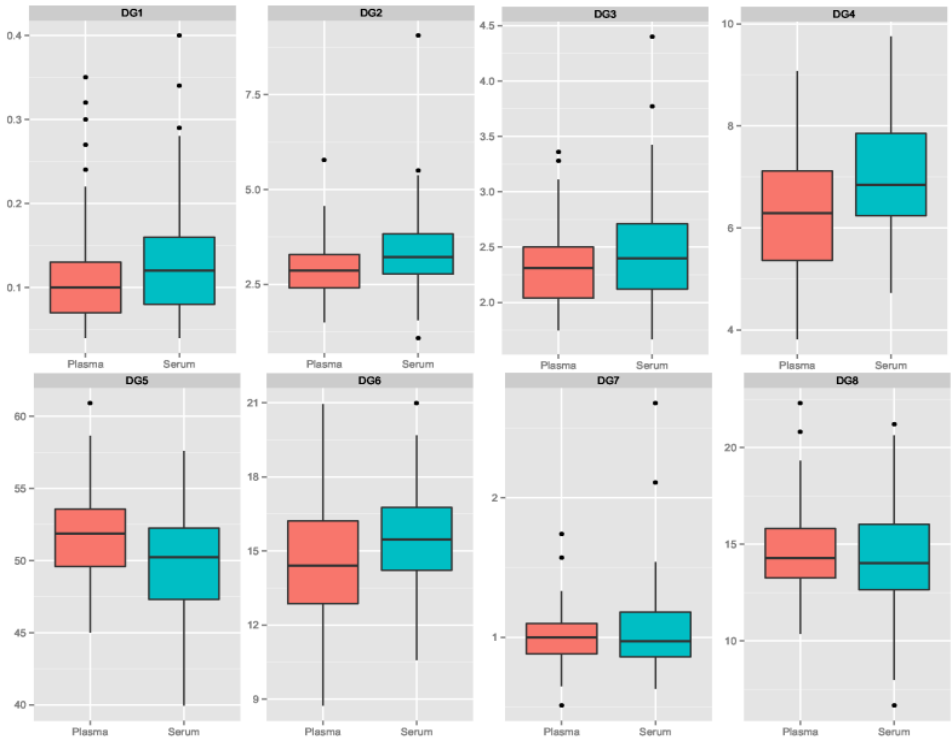


Figure 18: Box plots comparing plasma versus serum glycan values in the HNF1A-MODY cohort. Values for the peaks DG1-DG8.

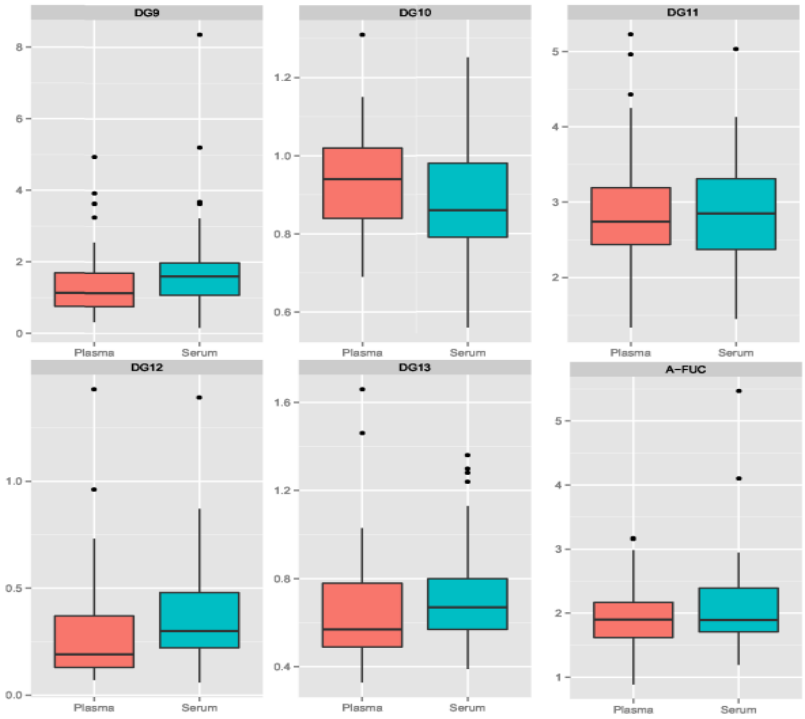


Figure 19: Box plots comparing plasma versus serum glycan values in the HNF1A-MODY cohort. Values for the peaks DG1-DG13 and calculated glycan feature A-Fuc (antennary fucosylation).

Table 21: P values calculated using Mann-Whitney U tests. Two groups of HNF1A-MODY samples, serum (n=49) and plasma (n=61), were tested for statistically significant difference. Statistically significant p value of 0.05 was corrected according to Bonferroni correction to 0.0017 (p*1/n); n=number of tested variables; Statistically significant values are highlighted in light red.

Glycan peak	Plasma vs. Serum p value	Glycan peak	Plasma vs. Serum p value
GP1	<0.001	**DG1**	0.039
GP2	0.007	**DG2**	0.023
GP3	<0.001	**DG3**	0.042
GP4	0.028	**DG4**	0.002
GP5	0.209	**DG5**	0.007
GP6	0.065	**DG6**	0.072
GP7	<0.001	**DG7**	0.817
GP8	0.189	**DG8**	0.243
GP9	0.121	**DG9**	0.024
GP10	0.833	**DG10**	0.008
GP11	<0.001	**DG11**	0.864
GP12	0.386	**DG12**	0.025
GP13	0.09	**DG13**	0.035
GP14	0.019		
GP15	0.456		
GP16	0.007		

4.5 COMPARISON OF DIFFERENT DIABETES COHORTS

All analyzed diabetes cohorts (GCK-MODY, HNF1A-MODY, HNF4A-MODY, Type 1 diabetes and Type 2 diabetes) are graphically compared in box plots (Figure 20-23). Glycan structures that gave the best differentiation (highest p values) between HNF1A-MODY and four other diabetes types and subtypes are displayed separately in Figure 23. Four diabetes cohorts were compared to HNF1A-MODY by nonparametric Mann-Whitney U test since data are not normally distributed. P values were calculated for all 29 measured glycan features to determine statistically significant difference between HNF1A-MODY and other diabetes cohorts. Level of the p value for the significance was corrected according to Bonferroni correction ($1/116*0.05$; $p < 4x10^{-4}$). Results for the Mann-Whitney statistic are shown in Table 22

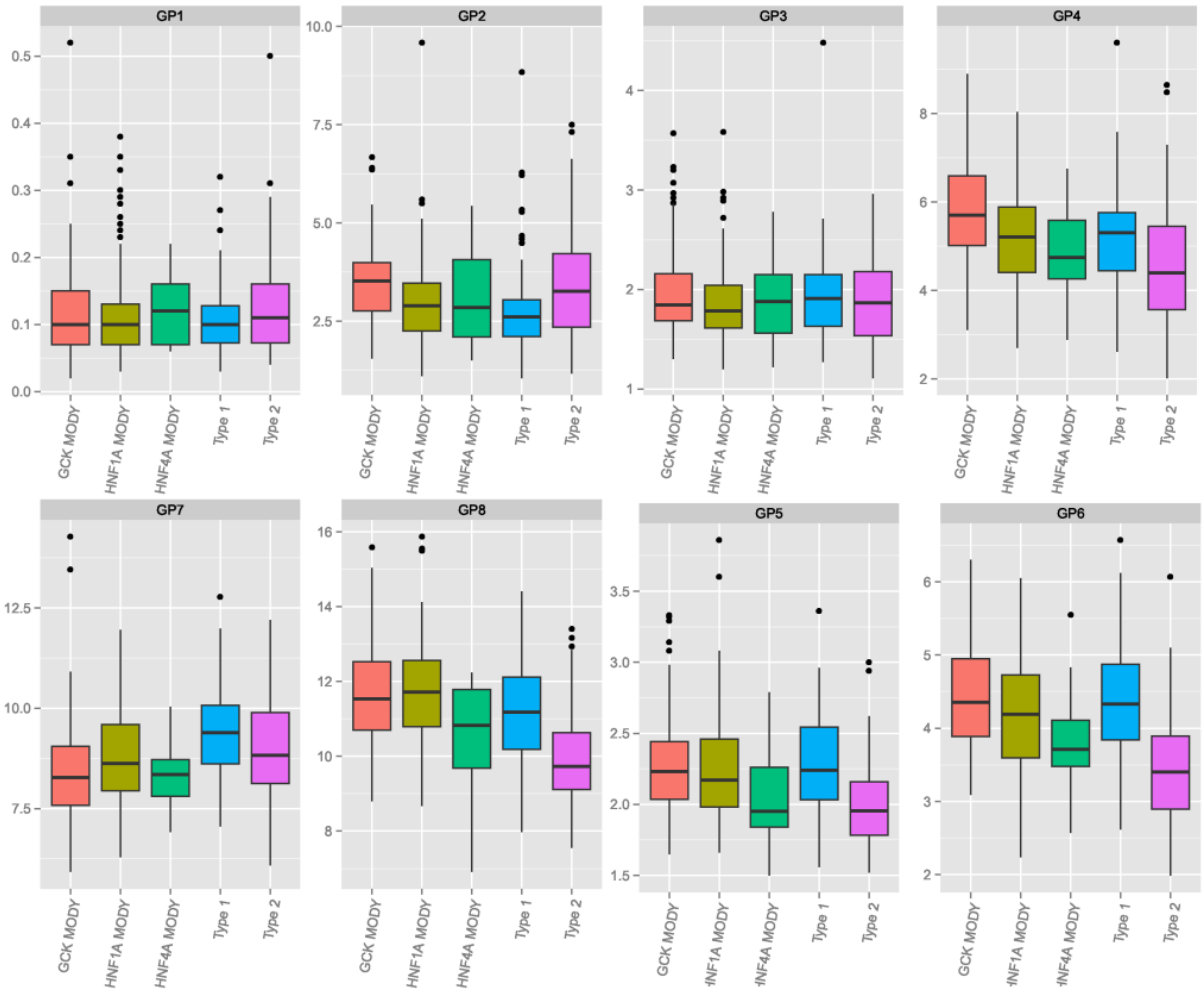


Figure 20: Box plots for five analyzed diabetes cohorts. GP1-GP8 - chromatography peaks of the initial glyco profile; light red-GCK-MODY; olive green-HNF1A-MODY; green-HNF4A-MODY; blue-Type 1 diabetes; violet-Type 2 diabetes.

Table 22: Mann-Whitney U test p values for all glycan traits comparing HNF1A-MODY patients and all other diabetes cohorts. Values lower than significance level are highlighted ($p < 4x10^{-4}$). P value 0.05 was corrected according to Bonferroni correction (1/116*0.05)

Glycan	vs. GCK-MODY	vs. HNF4A-MODY	vs. Type 1 diabetes	vs. Type 2 diabetes
GP1	4.762E-01	2.260E-01	8.036E-01	2.166E-01
GP2	3.766E-05	6.581E-01	8.632E-02	8.618E-02
GP3	8.097E-02	9.802E-01	2.464E-01	8.605E-01
GP4	2.935E-04	2.192E-01	9.031E-01	2.310E-04
GP5	3.596E-01	3.023E-02	2.188E-01	1.687E-06
GP6	6.857E-02	4.781E-02	4.063E-01	9.350E-09
GP7	2.320E-02	1.567E-01	1.113E-04	1.812E-01
GP8	6.911E-01	6.410E-03	1.974E-02	2.487E-11
GP9	3.185E-01	5.808E-01	5.334E-03	2.733E-03
GP10	6.010E-01	2.325E-02	7.410E-07	2.790E-02
GP11	6.507E-06	7.366E-01	1.181E-04	1.943E-03
GP12	3.378E-14	5.472E-01	9.910E-05	4.211E-02
GP13	3.062E-15	3.096E-01	3.261E-08	2.033E-08
GP14	1.180E-12	1.447E-06	1.048E-15	2.433E-17
GP15	1.447E-14	5.519E-01	9.191E-03	7.424E-15
GP16	2.445E-02	4.083E-03	6.973E-02	3.124E-06
DG1	7.869E-03	3.065E-01	8.346E-01	2.996E-04
DG2	3.855E-05	9.520E-01	1.022E-01	1.868E-01
DG3	9.686E-02	1.854E-01	3.227E-01	2.988E-01
DG4	1.770E-04	2.663E-01	6.791E-01	1.278E-03
DG5	3.552E-03	6.358E-02	2.545E-04	2.787E-01
DG6	9.496E-02	3.062E-01	4.849E-01	4.113E-05
DG7	1.585E-20	3.329E-06	1.526E-22	2.159E-22
DG8	1.875E-18	2.678E-01	3.665E-08	3.901E-05
DG9	2.824E-21	5.355E-07	2.053E-22	7.532E-19
DG10	4.677E-06	8.085E-05	1.534E-09	1.406E-19
DG11	1.640E-17	2.724E-01	2.134E-03	1.709E-05
DG12	2.248E-18	9.350E-07	3.522E-20	3.370E-15
DG13	7.317E-10	2.353E-03	1.902E-04	9.961E-06
C-FUC	3.467E-02	8.070E-01	1.245E-01	2.960E-03
A-FUC	2.662E-20	3.519E-06	1.701E-22	1.846E-22
DG7index	1.085E-15	1.636E-05	6.892E-19	2.064E-22
DG9index	2.683E-23	1.680E-06	8.086E-22	5.643E-18

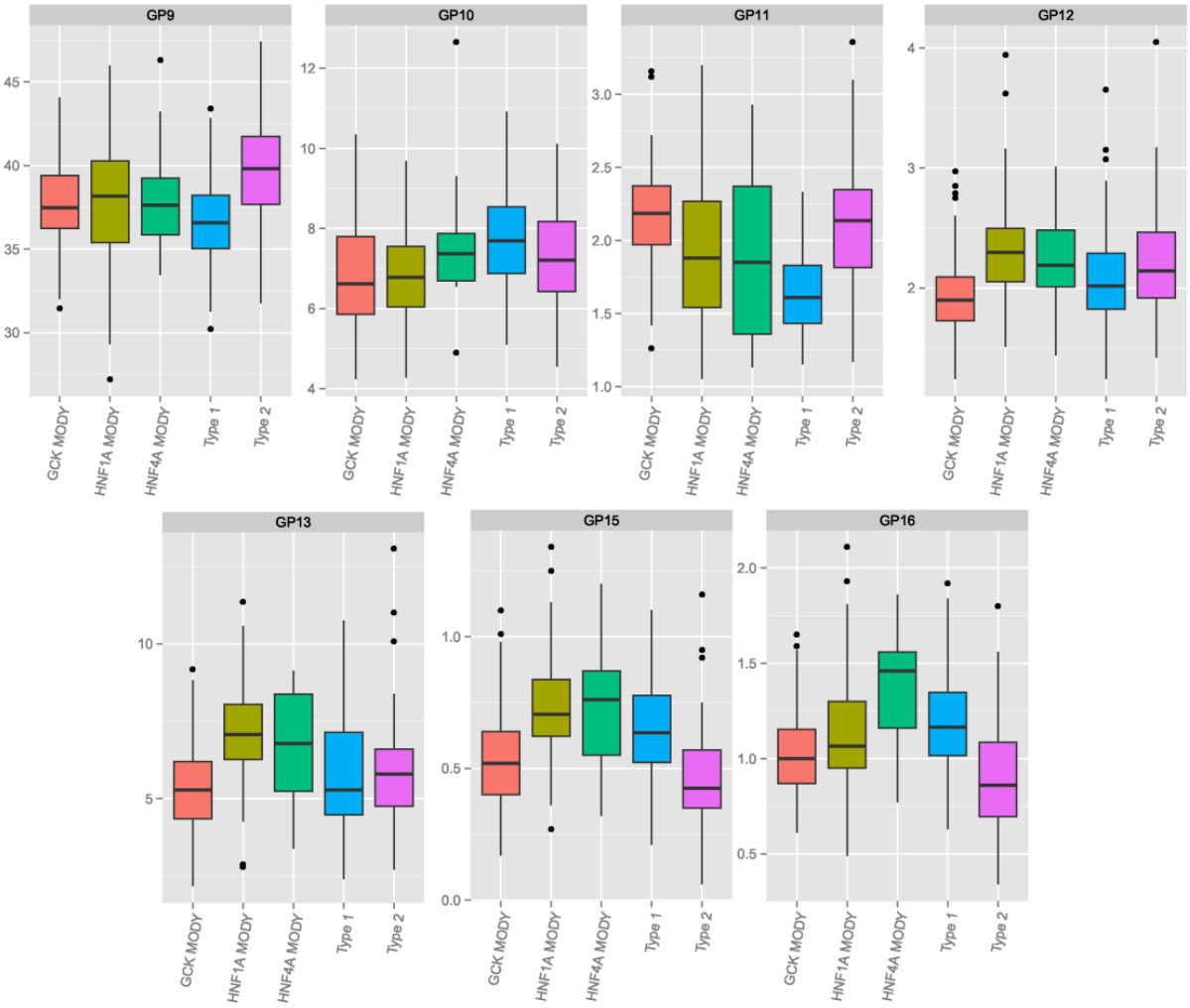


Figure 21: Box plots for five analyzed diabetes cohorts. GP9-GP13, GP15-GP16 - chromatography peaks of the initial glyco profile; light red-GCK-MODY; olive green-HNF1A-MODY; green-HNF4A-MODY; blue-Type 1 diabetes; violet-Type 2 diabetes.

In genome wide association study, glycan peaks DG11 and DG7 showed genome wide significance with two SNPs that were located in the HNF1A gene area. The effect sizes between those two peaks have opposite directions what is in accordance with the results obtained after the diabetes patients plasma glycome analysis. Value of DG7 peak, which contains only glycan structure with antennary fucose, is significantly lower in HNF1A-MODY patients than in general population cohort ($p<0.001$). On the other hand, DG11 value for HNF1A-MODY patients is higher in comparison to general population ($p<0.001$; Figure 24 and Figure 25). The latter glycan peak contains core fucosylated glycan structures without antennary fucosylation (Figure 26).

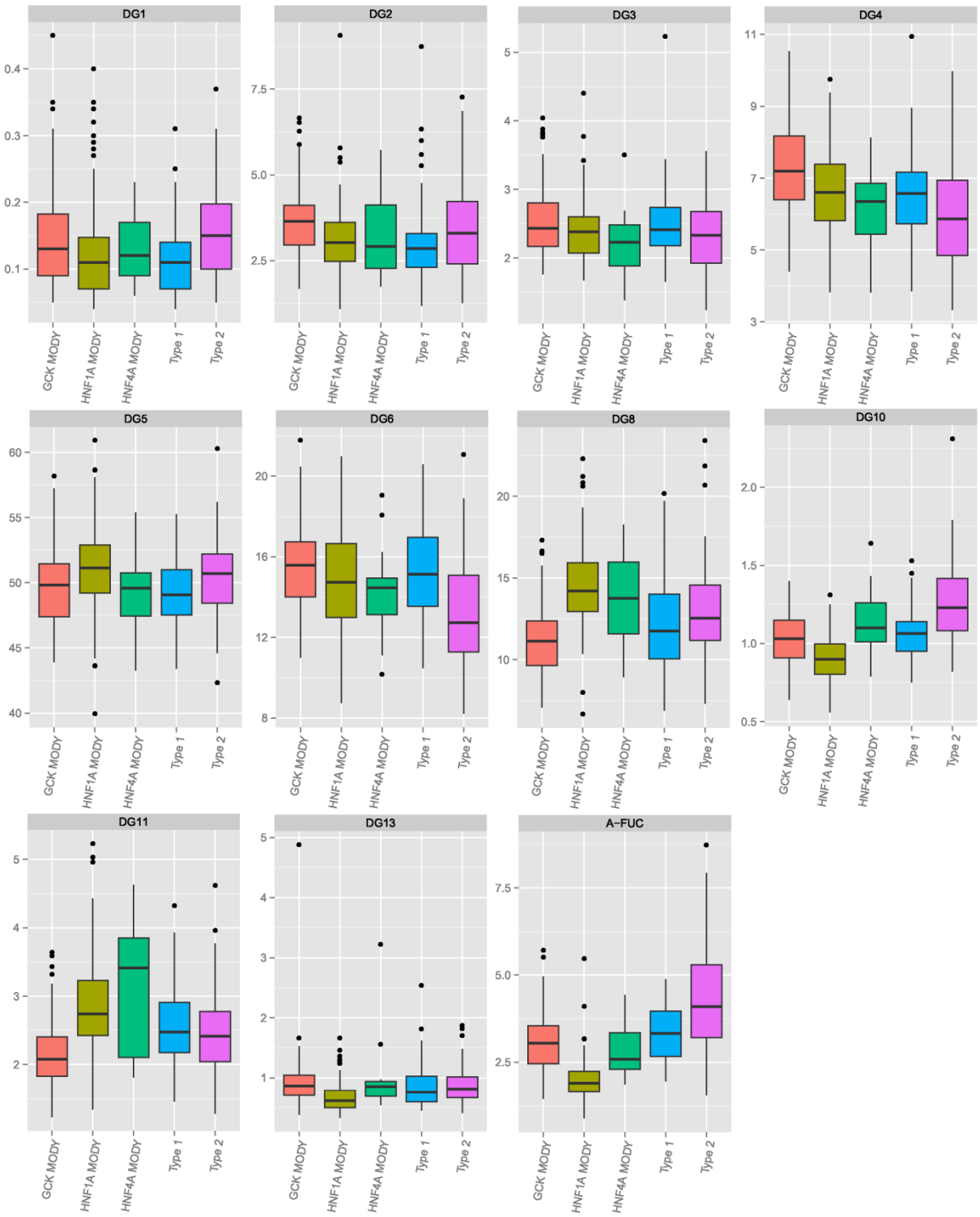


Figure 22: Box plots for five analyzed diabetes cohorts. DG1-DG6, DG8, DG10, DG11, DG13 - chromatography peaks of the desialylated glyco profile; Fuc-A – calculated glycan feature representing the portion of glycan structures with antennary Fucose; light red-GCK-MODY; olive green-HNF1A-MODY; green-HNF4A-MODY; blue-Type 1 diabetes; violet-Type 2 diabetes.

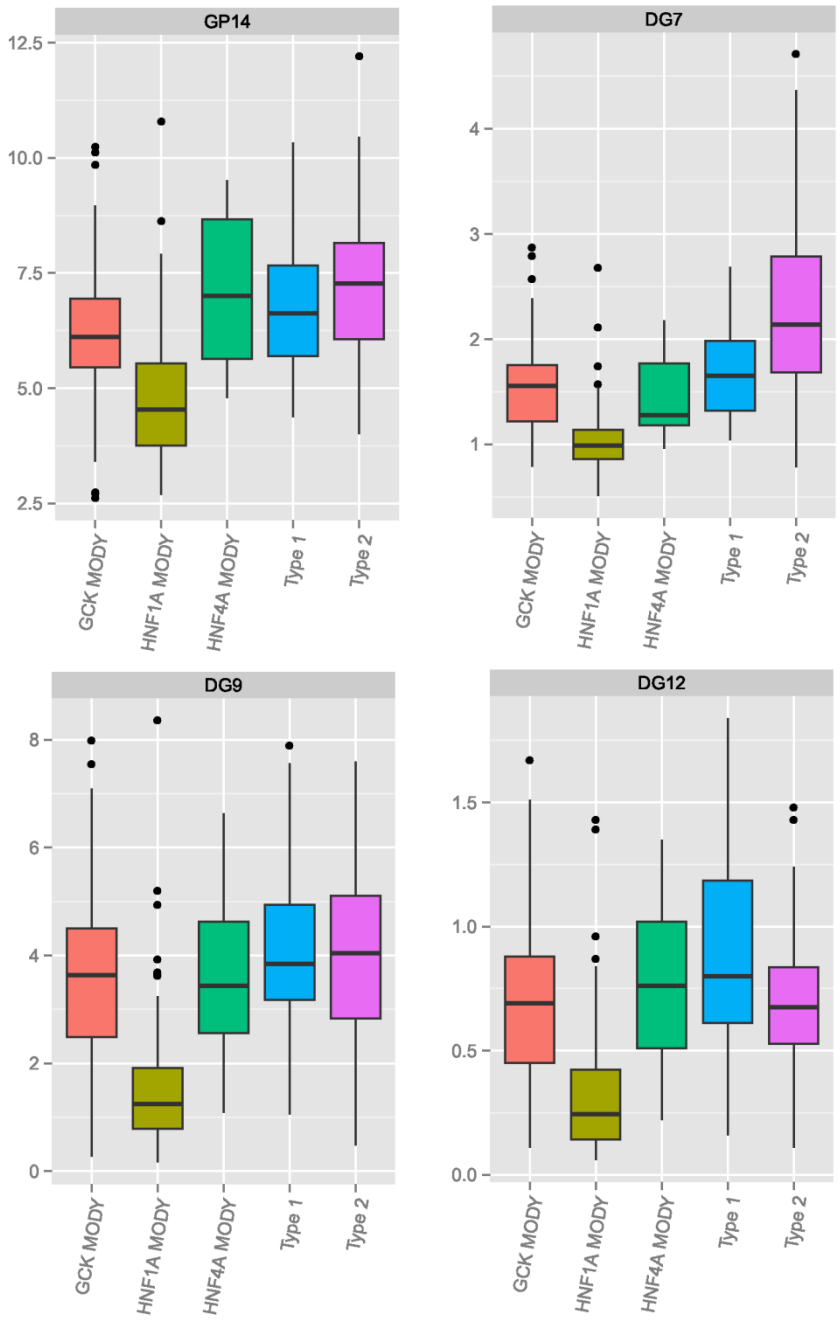


Figure 23: Box plots comparing five analyzed diabetes cohorts for four glycan features that displayed highest p values for Mann-Whitney U statistic. GP9; light red-GCK-MODY; olive green-HNF1A-MODY; green-HNF4A-MODY; blue-Type 1 diabetes; violet-Type 2 diabetes.

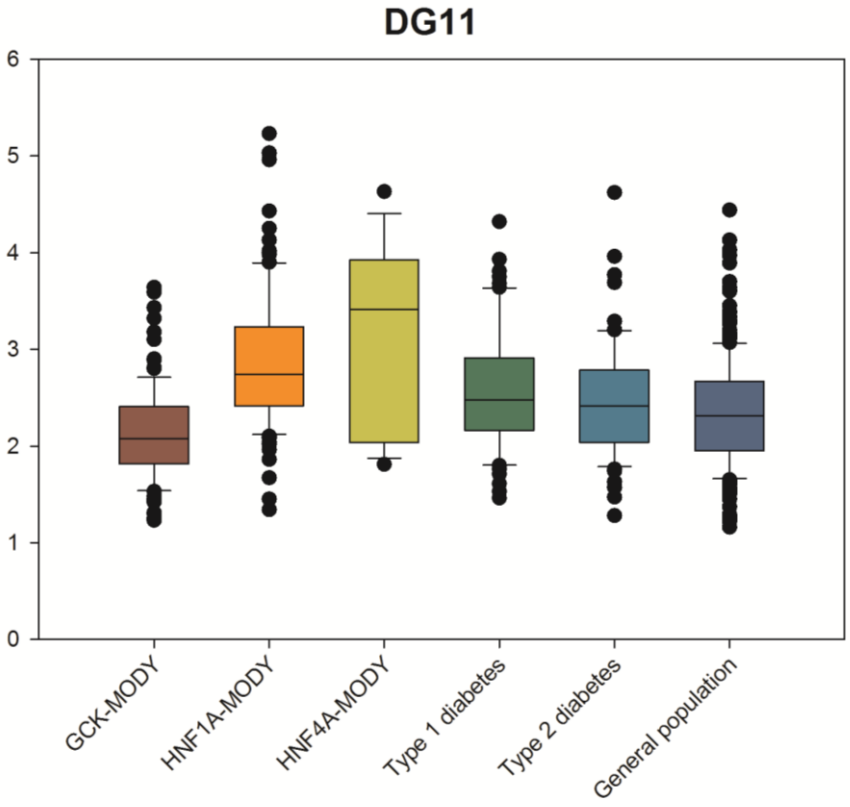


Figure 24: DG11 glycan peak versus all diabetes cohorts and general population

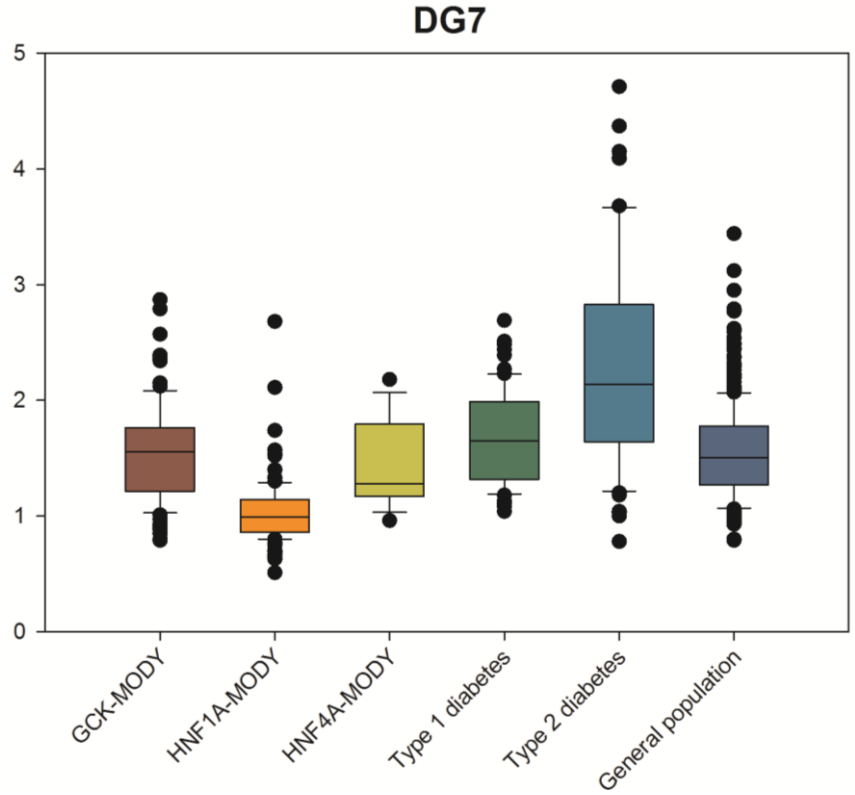


Figure 25: DG7 glycan peak versus all diabetes cohorts and general population

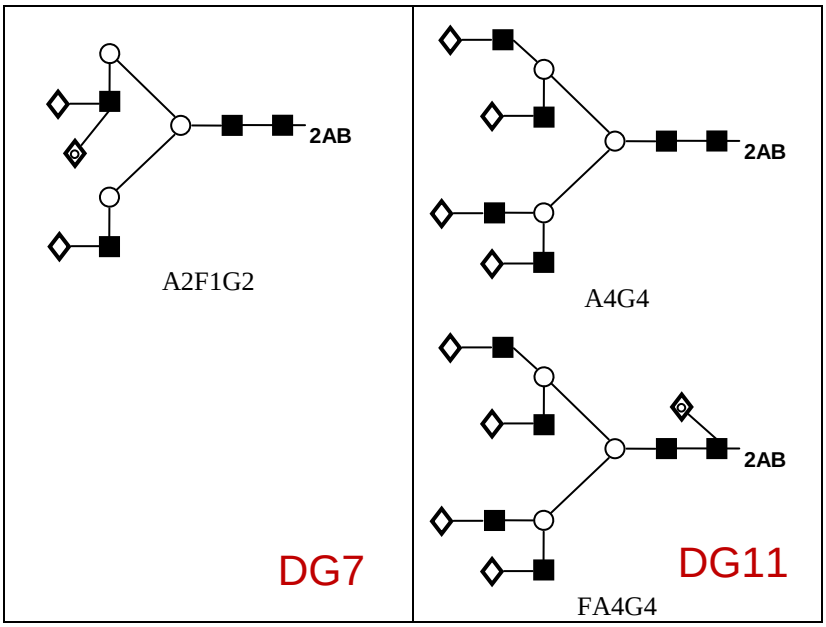


Figure 26: Glycan structures contained in the peaks DG7 and DG11. DG7 glycan peak contains only structures with antennary fucosylation, whilst DG11 peak contains only core fucosylated structures. . ■ - N-acetylglucosamine, ○ - mannose, ◇ - galactose, ◈ - fucose; 2AB – 2-aminobenzamide.

Additionally, glycan peaks GP14, DG9 and DG12 displayed highly lower values for HNF1A-MODY cohort in comparison to four other diabetes groups. All of them contain structures with antennary fucosylation. Difference between DG9 and DG7 glycan structures is in the number of glycan branches. DG7 contains biantennary whilst DG9 contains triantennary glycans.

Following the findings of genome wide association study, where two hits showed effects in opposite directions (anntenary vs core fucosylation), we introduced two additional values for diabetes cohorts discrimination called DG9 index and DG7 index. They represent DG9 to (DG9 + DG8) ratio and DG7 to (DG6 + DG5) ratio respectively. DG9 index measures the proportion of triantennary glycans that are antennary fucosylated whilst DG7 index summarizes the proportion of biantennary fucosylated glycans. Both indexes displayed improved discrimination (Table 22). DG7 index, in contrary to DG9 index, includes extremely low values approaching zero what results in highly tailed distribution, an effect which can cause problems in patient stratification (Figure 27).

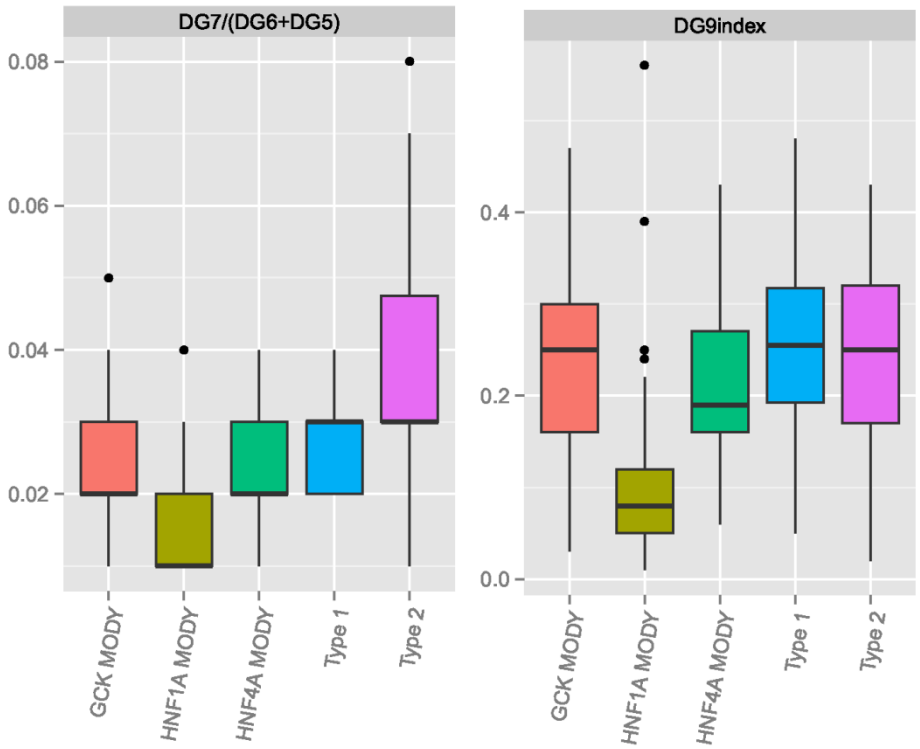


Figure 27: Box plots for DG7 and DG9 indexes. light red-GCK-MODY; olive green-HNF1A-MODY; green-HNF4A-MODY; blue-Type 1 diabetes; violet-Type 2 diabetes.

Median (interquartile range) DG9-glycan index levels were substantially lower in subjects with HNF1A-MODY [0.08 (0.06–0.11)] than in those with young-onset type 2 diabetes [0.25 (0.17–0.32); P<0.0004 vs. HNF1A-MODY], type 1 diabetes [0.25 (0.24–0.26); P<0.0004 vs. HNF1A-MODY], or GCK-MODY [0.25 (0.20–0.27); P<0.0004 vs. HNF1A-MODY]. Cases of HNF4A-MODY showed DG9-glycan index levels between those inHNF1A-MODY and other forms of diabetes [0.20 (0.18–0.22); P<0.0004 vs.HNF1A-MODY]. Adjustment for significant covariates (age at sampling, gender, BMI) had no appreciable impact on the magnitude or significance of differences in DG9-glycan index values between groups.

4.6 Receiver operating characteristic (ROC) curve analyses.

To test whether glycan profiling had potential as a clinically valid screening test, area under curve (AUC) measures of discriminative accuracy were derived from ROC curve analyses (Figure 28 – 31; Table 23). The DG9-index AUC value was 0.92 for HNF1A-MODY against type 1 diabetes and 0.89 for HNF1A-MODY against early-onset type 2 diabetes. Similar discrimination was observed for the comparison of HNF1A-MODY and GCK-MODY (AUC = 0.89), but the DG9-glycan index performed less well in differentiating HNF4A-MODY and HNF1A-MODY (AUC = 0.86).

Table 23: Area under curve (AUC) values derived from receiver operating characteristic (ROC) curve analysis describing discriminatory accuracy for HNF1A-MODY against GCK-MODY, HNF4A-MODY, type 1 diabetes and type 2 diabetes.

	vs. GCK-MODY	**vs. HNF4A-MODY**	**vs. Type 1 diabetes**	**vs. Type 2 diabetes**
GP14	0.78	0.86	0.86	0.88
DG7	0.86	0.85	0.93	0.94
DG9	0.87	0.88	0.93	0.9
DG7-index	0.86	0.86	0.94	0.95
DG9-index	0.89	0.86	0.92	0.89

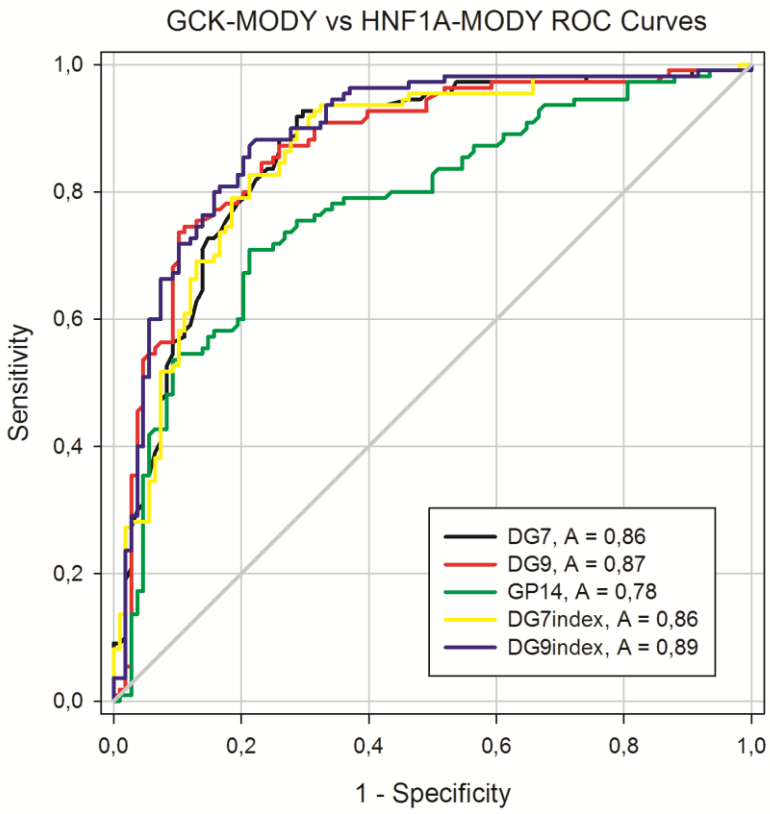


Figure 28: ROC curve illustrating the performance of the DG7, DG9 and GP14 glycan peaks as well as DG9 – and DG7-index to discriminate HNF1A-MODY and GCK-MODY.

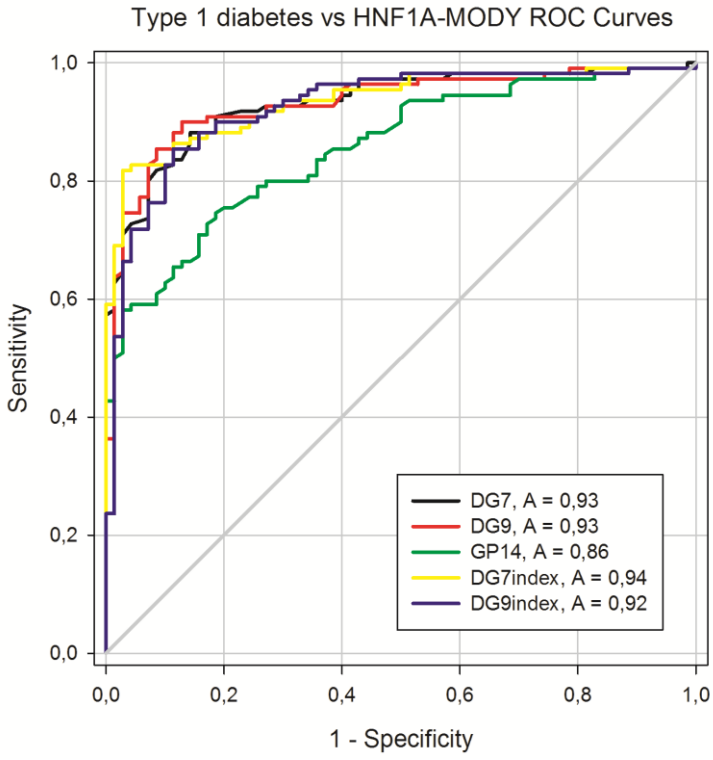


Figure 29: ROC curve illustrating the performance of the DG7, DG9 and GP14 glycan peaks as well as DG9 – and DG7-index to discriminate HNF1A-MODY and Type 1 diabetes.

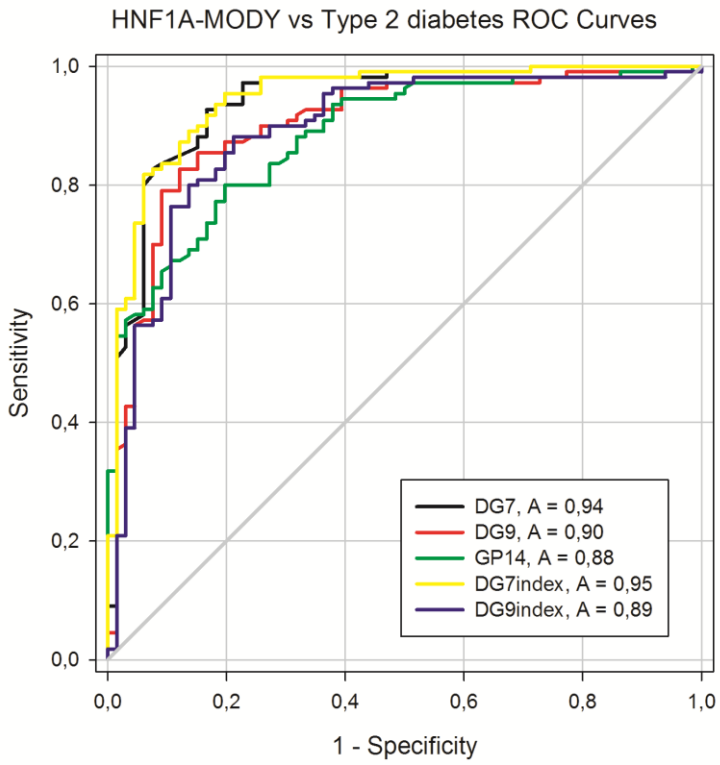


Figure 30: ROC curve illustrating the performance of the DG7, DG9 and GP14 glycan peaks as well as DG9 – and DG7-index to discriminate HNF1A-MODY and type 2 diabetes.

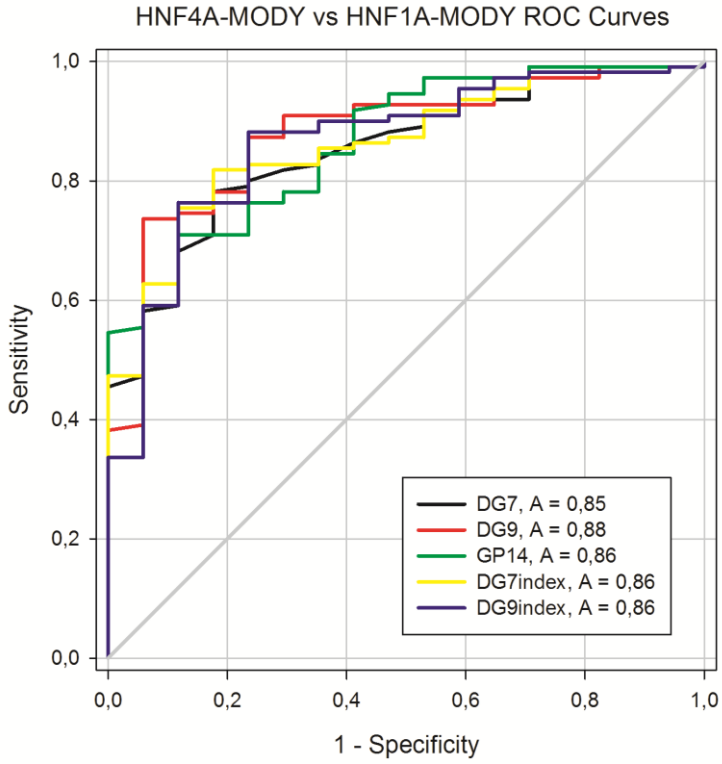


Figure 31: ROC curve illustrating the performance of the DG7, DG9 and GP14 glycan peaks as well as DG9 – and DG7-index to discriminate HNF1A-MODY and HNF4A-MODY.

5 DISCUSSION

5.1 ROBUSTNESS OF THE GLYCAN ANALYSIS PROCESS

High throughput process of the glycan analysis was evaluated using Design of Experiment approach. Eleven factors were selected for the testing of the effects they have on the method robustness and stability of the end results. Absolute values of the factor responses might not represent values which could extremely affect the glycome analysis, but it should be only considered for the range defined with factors high and low levels. Factor values outside, in this work defined, boundaries are unknown area and might cause nonlinear changes and bad results. Observed effects can provide, even after performing only initial screening experiments, a conclusion about favorable factor levels (low or high level) which can significantly improve our laboratory procedure.

Glycan release, labeling and HPLC analysis of each sample, and therefore of each glycan structure present in a plasma sample, are performed in a batch and therefore there is no possibility of exposing different glycan groups to different conditions during this process. All significant factor effects are having the same direction (Figure 13Figure 14) and therefore enable us to set the factors at the lower or upper level and provide, even at the screening level, the information about critical steps in the procedure.

Design of experiment techniques are a powerful tool for describing of different processes. Being aware of the critical steps in the process is of the extreme value especially in improving the method and in elimination of errors and deviations. Classic “one at the time” approach would mean an enormous number of experiments for testing a high number of factors, as in the case of this work, eleven factors. To save money and time and still be able to obtain quality and informative data, experimental designs as Plackett-Burman can provide an ideal compromise[66].

Results obtained showed that 8 out of 11 tested process parameters showed statistically significant effect for the 6 response values observed. Recorded CV values, under all of the investigated operating conditions, showed small increase regarding previously recorded values[9] meaning that the glycan analysis process is robust.

In the course of this work several critical steps in the glycan analysis process were indentified which should be carried out strictly according to SOP since the deviation from defined condition causes higher variability in the glycan analysis. Time of exposure to DTT (X1) should be set to lower level (15 minutes) since prolonged protein reduction by DTT results in higher variability of the glycan peak area percentage. The gel block encompassing plasma proteins shouldn't be frozen prior to the transfer to filter plates since freezing displayed an impact in the variability for GP8. Enzyme concentration can be lowered and it will have no impact on the variability. Digestion should continue for more than 16 hours and formic acid treatment shouldn't last more than 40 minutes since prolonged treatment causes greater variation. Glycan labeling procedure should last for 2 hours and post labeling acetonitrile wash should continue for 15 minutes. After the glycan preparation steps, the HPLC analysis of labeled glycans should be carried out as soon as possible.

Next step in the method improvement could be full factorial design testing only effects which were identified as important in this study, allowing us to estimate all main effects and two-factor interactions aliased only by three-factor or higher order interactions.

5.2 HNF1A TRANSCRIPTION FACTOR

Results presented in this work imply that HNF1A transcriptionally regulates the expression of genes involved in the steps of fucosylation what is additionally supported by the functional studies[27], and the fact that a functionally related transcription factor, HNF4A was previously shown to bind the regulatory elements of the GDP-mannose-4,6-dehydratase (GMDS) gene in a genome-wide ChIP-ChIP. GMDS is involved in the de novo pathway of L-fucose synthesis to produce GDP-fucose, the substrate used by both core and antennary fucosyltransferases to N-glycosylated proteins[53]. HNF4A directly regulates the expression of the hepatic fucosyl-transferase VI gene (FUT6)[75] whilst earlier genome wide studies showed connection between HNF4A and FUT2 and FUT5 (genes for fucosyltransferase) in intestinal epithelial cells[81].

By performing the first genome-wide association analysis (GWAS) of protein glycosylation first step towards the mapping of the complex network of genes that

regulate protein N-glycosylation is made. A gene encoding the transcription factor HNF1A, with previously unknown biological links to glycosylation, is shown to be strongly associated with the relative proportions of fucosylated plasma N-glycans. Both core and antennary fucosylation displayed significant effects in opposite direction. As shown in Figure 23, patients with mutations in HNF1A gene have significantly lower levels of antennary fucosylation and significantly higher levels of core fucosylated structures than general population ($p<0.001$).

The possible function(s) of HNF1A are a focus of intense current interest following its recently reported associations in GWAS with plasma C-reactive protein (CRP)[76], gamma-glutamyl transferase (GGT)[77], LDL cholesterol and apolipoprotein[78] and coronary artery disease[79].

Additional analysis of gene knockdowns (RNAi) showed that HNF1A is an upstream regulator of several key genes involved in different stages of the fucosylation pathway. HNF1A binds the promoters of two genes, fucokinase and GMDS, required for de novo and salvage pathways of fucose synthesis[27]. Fucose synthesis is the rate limiting step for fucosylation in eukaryotes and prokaryotes[80] and, by upregulating its synthesis, HNF1A increases the availability of fucose to the glycosylation machinery. HNF1A reciprocally regulates core versus antennary fucosylation; while activating different fucosyltransferases (FUTs) involved in antennary fucosylation[27]. The reciprocity of this regulation is displayed in the levels of peaks DG7 and DG11, containing antennary fucosylated and core fucosylated structures respectively.

5.3 Fucosylation as HNF1A-MODY clinical biomarker

The study confirms the hypothesis that the glycan profile of plasma proteins is altered substantially in patients with HNF1A mutations. It was also demonstrated that these differences could be exploited as biomarkers in diabetes diagnostics and showed that several glycan peaks, GP14, DG7, DG9 and DG12 as well as two calculated indexes, DG9-glycan index and DG7-glycan index, can discriminate HNF1A-MODY from GCK-MODY, HNF4A-MODY, type 1 and type 2 diabetes. Several efforts to improve diagnostic performance by identifying biochemical markers specific for MODY subtypes showed unsatisfactory effects and demonstrate the importance of the correct diagnosis for monogenic diabetes subtypes[82–87]. Recently it was

demonstrated that individuals with HNF1A-MODY have low levels of C-reactive protein and that hs-CRP (high sensitivity-CRP) assays can discriminate well between HNF1A-MODY and both type 2 diabetes (AUC 0.94) and HNF4A-MODY[54,88,89]. However, hs-CRP does not provide good discrimination between HNF1A-MODY and type 1 diabetes. DG9-glycan index provided better discrimination compared with hs-CRP (AUC 0.94 vs. 0.83 for hs-CRP). Furthermore, hs-CRP is an acute inflammatory marker, and diagnostic discrimination can be disturbed by intercurrent infection[90]. Potential advantages of the glycan index in this context include stability over time[91] and differentiation of HNF1A-MODY from both common types of diabetes. This suggests that the DG9-glycan index is less prone to spurious elevation from intercurrent infection than hs-CRP, although this will require confirmation in larger numbers[90].

The ability to discriminate between HNF1A-MODY and type 1 diabetes in subjects with recently diagnosed diabetes is particularly important because diagnostic mis-classification can lead to the unwarranted decision to recommend lifelong therapy with exogenous insulin. Detectable C-peptide can indicate HNF1A-MODY rather than type 1 diabetes of long duration, but it is not helpful close to a diagnosis of diabetes because a substantial proportion of type 1 diabetes patients retain some production of endogenous insulin[92].

Since glycan profiles are stable within an individual over time, which suggests these measures will continue to provide useful discriminative power from type 1 diabetes close to diagnosis and would not be hindered by the occurrence of honeymoon period[9,91]. The addition of the DG9-glycan index to existing biomarkers such as hs-CRP[82], 1,5-anhydroglucitol[93], pancreatic autoanti-bodies[94], and C-peptide[92] should improve the capacity for clinical discrimination of all diabetes subtypes.

Further validation of the DG9-glycan index in unselected groups of subjects with young-onset diabetes will be required to assess performance in a more typical clinical scenario in which extensive prior screening for monogenic disease has not been undertaken. The use of the DG9-glycan index in clinical practice currently is restricted by the cost and limited availability of accurate glycan profiling. Clinical translation is, therefore, dependent on the implementation of a focused assay for specific glycan moieties rather than the global chromatographic profiling used in this study. The

DG9-glycan index seems to be the most promising candidate from this study, although the validation study suggests that other measures (such as the DG7-to-(DG5+DG6) ratio) might have superior performance when plasma (rather than serum) samples are available.

Glycan antennary fucosylation analysis represents a potential adjunct in the diagnosis algorithm of HNF1A-MODY. Due to its relative low-cost in comparison to expensive diagnostic sequencing it could play a significant role in prescreening process of patients with high risk of HNF1A-MODY.

Beside the diagnostic itself, biomarkers like DG9 index have potential in determining the significance of numerous genetic variances that have been documented in MODY patients. In this moment there are 193 documented mutations that affect the function of HNF1A and the possibility of DG9 index to discriminate between them has been already demonstrated[90]. With anticipated reduction in cost of diagnostic sequencing the importance of glycan index will be shifting towards accurate interpretation of novel uncovered variants[90].

6 CONCLUSION

Screening experiment provided quality data which resolved the questions about better conditions and robustness of the high throughput glycan analysis. Fractional factorial design used in this study enabled us to test great deal of critical steps in time, labor and money saving manner. By identifying those critical steps we improved our method and in that way obtained more accurate and reliable data for studies of large cohorts

Antennary fucosylation of plasma glycans is significantly lower in patients with mutations in gene for transcription factor HNF1A and core fucosylation is significantly increased in comparison to general population. Those results are confirming initial findings of genome wide association studies that showed genome wide significance in the opposite direction for HNF1A genetic markers and antennary and core fucosylation.

Several glycan peaks containing antennary fucosylated structures displayed significantly lower values in HNF1A-MODY cohort and represent a pool of potential indexes as biomarkers for this type of MODY (maturity-onset diabetes of the young), and for biochemical activity of HNF1A transcription factor.

The DG9-glycan index, as a biomarker for HNF1A-MODY, demonstrated the best overall result although all glycan peaks with significantly lower level of antennary fucosylated structures should be further examined using larger cohorts of potential misdiagnosed patients. The DG9-glycan index is a compelling example of the potential for rapid clinical translation of a genetic discovery originating from a genome wide association analysis.

7 REFERENCES

1. Apweiler, R., Hermjakob, H. & Sharon, N. On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochimica et Biophysica Acta* **1473**, 4–8 (1999).

2. Hart, G. W., Housley, M. P. & Slawson, C. Cycling of O-linked β-N-acetylglucosamine on nucleocytoplasmic proteins. *Nature* **446**, 1017–1022 (2007).

3. Wells, L., Vosseller, K. & Hart, G. W. Glycosylation of Nucleocytoplasmic Proteins: Signal Transduction and O-GlcNAc. *Science* **291**, 2376–2378 (2001).

4. Dennis, J. W., Lau, K. S., Demetriou, M. & Nabi, I. R. Adaptive regulation at the cell surface by N-glycosylation. *Traffic (Copenhagen, Denmark)* **10**, 1569–78 (2009).

5. Marth, J. D. & Grewal, P. K. Mammalian glycosylation in immunity. *Nat Rev Immunol.* **8**, 874–887 (2009).

6. Varki, A. *et al. Essentials of Glycobiology*. (Cold Spring Harbor Laboratory Press: New York, 1999).

7. Sharon, N. & Lis, H. Lectins as cell recognition molecules. *Science* **246**, 227–234 (1989).

8. Gornika, O., Pavić, T. & Lauc, G. Alternative glycosylation modulates function of IgG and other proteins — Implications on evolution and disease. *Biochimica et Biophysica Acta (BBA) - General Subjects* **1820**, 1318–1326 (2011).

9. Knezevic, A. *et al.* Variability , Heritability and Environmental Determinants of Human Plasma N-Glycome. *Journal of Proteome Research* **7**, 694–701 (2008).

10. Zoldoš, V., Novokmet, M., Bečeheli, I. & Lauc, G. Genomics and epigenomics of the human glycome. *Glycoconjugate Journal* **30**, 41–50 (2013).

11. Marek, K. W., Vijay, I. K. & Marth, J. D. A recessive deletion in the GlcNAc-1-phosphotransferase gene results in peri-implantation embryonic lethality. *Glycobiology* **9**, 1263–71 (1999).

12. Yoshida, A. Identification of genotype of blood group A and B. *Blood* **55**, 119–23 (1980).

13. Dhiman, N. *et al.* genes and immunity to rubella vaccine. *Immunogenetics* **62**, 197–210 (2010).

14. Ohtsubo, K. & Marth, J. D. Glycosylation in cellular mechanisms of health and disease. *Cell* **126**, 855–867 (2006).

15. Cummings, R. D. The repertoire of glycan determinants in the human glycome. *Molecular bioSystems* **5**, 1087–104 (2009).

16. Foulquier, F. *et al.* Conserved oligomeric Golgi complex subunit 1 deficiency reveals a previously uncharacterized congenital disorder of glycosylation type II. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 3764–9 (2006).

17. Wu, X. *et al.* Mutation of the COG complex subunit gene COG7 causes a lethal congenital disorder. *Nature medicine* **10**, 518–23 (2004).

18. Jaeken, J. & Matthijs, G. Congenital Disorders of Glycosylation: A Rapidly Expanding Disease Family. *Annual Review of Genomics and Human Genetics* **8**, 261–278 (2007).

19. Klein, A. Human total serum N-glycome. *Advances in Clinical Chemistry* **46**, 51–85 (2008).

20. Kaneko, Y., Nimmerjahn, F. & Ravetch, J. V Anti-inflammatory activity of immunoglobulin G resulting from Fc sialylation. *Science* **313**, 670–673 (2006).

21. Masuda, K. *et al.* Enhanced binding affinity for FcgammaRIIIa of fucose-negative antibody is sufficient to induce maximal antibody-dependent cellular cytotoxicity. *Molecular Immunology* **44**, 3122–3131 (2007).

22. Nimmerjahn, F. & Ravetch, J. V Divergent immunoglobulin g subclass activity through selective Fc receptor binding. *Science* **310**, 1510–1512 (2005).

23. Zaia, J. Mass Spectrometry and Glycomics. *Current Pharmaceutical Design* **14**, 401–418 (2010).

24. Zaia, J. *Mass spectrometric ionization of carbohydrates. In Encyclopedia of Mass Spectrometry*. (Elsevier, New York: New York, 2007).

25. Wada, Y. *et al.* Comparison of the methods for profiling glycoprotein glycans--HUPO Human Disease Glycomics/Proteome Initiative multi-institutional study. *Glycobiology* **17**, 411–22 (2007).

26. Alpert, A. J. Hydrophilic-interaction chromatography for the separation of peptides, nucleic acids and other polar compounds. *Journal of Chromatography A* **499**, 177–196 (1990).

27. Lauc, G. *et al.* Genomics Meets Glycomics—The First GWAS Study of Human N-Glycome Identifies HNF1α as a Master Regulator of Plasma Protein Fucosylation. *PLoS Genetics* **6**, e1001256 (2010).

28. Callewaert, N. *et al.* Noninvasive diagnosis of liver cirrhosis using DNA sequencer–based total serum protein glycomics. *Nature Medicine* **10**, 429–434 (2004).

29. Harvey, D. J. Matrix-assisted laser desorption/ionization mass spectrometry of carbohydrates and glycoconjugates. *International Journal of Mass Spectrometry* **226**, 1–35 (2003).

30. Royle, L. *et al.* HPLC-based analysis of serum N-glycans on a 96-well plate platform with dedicated database software. *Analytical biochemistry* **376**, 1–12 (2008).

31. Pucić, M. *et al.* High throughput isolation and glycosylation analysis of IgG-variability and heritability of the IgG glycome in three isolated human populations. *Molecular & cellular proteomics : MCP* **10**, M111.010090 (2011).

32. Rudd, P. M., Rudan, I. & Wright, A. F. High-Throughput Glycome Analysis Is Set To Join High-Throughput Genomics. *Journal of proteom research* **8**, 1105 (2009).

33. Bush, W. S. & Moore, J. H. Chapter 11: genome-wide association studies. *PLoS computational biology* **8**, e1002822 (2012).

34. Lin, J.-M. *et al.* Complement factor H variant increases the risk for early age-related macular degeneration. *Science* **28**, 1416–20 (2005).

35. Klein, R. J. *et al.* Complement Factor H Polymorphism in Age-Related Macular Degeneration. *Science* **308**, 385–389 (2005).

36. Genomes Project Consortium The 1000 A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–73 (2010).

37. Griffith, O. L. *et al.* ORegAnno: an open-access community-driven resource for regulatory annotation. *Nucleic acids research* **36**, D107–13 (2008).

38. Reich, D. E. & Lander, E. S. On the allelic spectrum of human disease. *Trends in genetics : TIG* **17**, 502–10 (2001).

39. Huffman, J. E. *et al.* Polymorphisms in B3GAT1, SLC9A9 and MGAT5 are associated with variation within the human plasma N-glycome of 3533 European adults. *Human molecular genetics* **20**, 5000–11 (2011).

40. Lebrilla, C. B. & An, H. J. The prospects of glycan biomarkers for the diagnosis of diseases. *Molecula biosystems* **5**, 17–20 (2009).

41. Wuhrer, M. Glycosylation profiling in clinical proteomics: heading for glycan biomarkers. *Expert Review of Proteomics* **4**, 135–136 (2007).

42. Vanderschaeghe, D. *et al.* High-throughput profiling of the serum N-glycome on capillary electrophoresis microfluidics systems: toward clinical implementation of GlycoHepatoTest. *Analytical chemistry* **82**, 7408–7415 (2010).

43. McCarthy, M. I. & Hattersley, A. T. Learning from molecular genetics: novel insights arising from the definition of genes for monogenic and type 2 diabetes. *Diabetes* **57**, 2889–98 (2008).

44. Froguel, P. *et al.* Close linkage of glucokinase locus on chromosome 7p to early-onset non-insulin-dependent diabetes mellitus. *Nature* **356**, 162–164 (1992).

45. Hattersley, A. *et al.* Linkage of type 2 diabetes to the glucokinase gene. *Lancet* **339**, 1307–1310 (1992).

46. Berg, J. M., Tymoczko, J. L. & Stryer, L. *Biochemistry*. (W H Freeman: New York, 2002).

47. Yamagata, K. *et al.* Mutations in the hepatocyte nuclear factor-4alpha gene in maturity-onset diabetes of the young (MODY1). *Nature* **384**, 458–460 (1996).

48. Yamagata, K. *et al.* Mutations in the hepatic nuclear factor 1 alpha gene in maturity-onset diabetes of the young (MODY3). *Nature* **384**, 455–458 (1996).

49. Stoffers, D., Ferrer, J., Clarke, W. & Habener, J. Early-onset type-II diabetes mellitus (MODY4) linked to IPF1. *Nature Genetics* **17**, 138–139 (1997).

50. Horikawa, Y. *et al.* Mutation in hepatocyte nuclear factor-1 beta gene (TCF2) associated with MODY. *Nature Genetics* **17**, 384–385 (1997).

51. Malecki, M. *et al.* Mutations in NEUROD1 are associated with the development of type 2 diabetes mellitus. *Nature Genetics* **23**, 323–328 (1999).

52. Fajnas, S. S., Bell, G. I. & Polonsky, K. S. Molecular mechanisms and clinical pathophysiology of maturity-onset diabetes of the young. *New england journal of medicine* **345**, 971–980 (2001).

53. Odom, D. T. *et al.* Control of Pancreas and Liver Gene Expression by HNF Transcription Factors. *Science* **303**, 1378–1381 (2004).

54. Ellard, S. & Colclough, K. Mutations in the genes encoding the transcription factors hepatocyte nuclear factor 1 alpha (HNF1A) and 4 alpha (HNF4A) in maturity-onset diabetes of the young. *Human Mutation* **27**, 854–869 (2006).

55. Shepherd, M. *et al.* No Deterioration in Glycemic Control in HNF-1α Maturity-Onset Diabetes of the Young Following Transfer From Long-Term Insulin to Sulphonylureas. *Diabetes care* **26**, 3191–3192 (2003).

56. Pearson, E. R. *et al.* Genetic cause of hyperglycaemia and response to treatment in diabetes. *Lancet* **362**, 1275–1281 (2003).

57. Pearson, E., Liddell, W., Shepherd, M., Corrall, R. & Hattersley, A. Sensitivity to sulphonylureas in patients with hepatocyte nuclear factor 1 alpha gene mutations: evidence for pharmacogenetics in diabetes. *Diabetic medicine* **17**, 543–545 (2000).

58. Seino, S. ATP-SENSITIVE POTASSIUM CHANNELS : A Model of Heteromultimeric Potassium Channel / Receptor Assemblies. *Annual Review of Physiology* **61**, 337–62 (1999).

59. Hathout, E. *et al.* A case of hepatocyte nuclear factor-1 alpha diabetes/MODY3 masquerading as type 1 diabetes in a Mexican-American adolescent and responsive to a low dose of sulphonylurea. *Diabetes Care* **22**, 867–868 (1999).

60. Murphy, R., Ellard, S. & Hattersley, A. T. Clinical implications of a molecular genetic classification of monogenic beta-cell diabetes. *Nature clinical practice Endocrinology metabolism* **4**, 200–213 (2008).

61. Wang, H., Antinozzi, P., Hagenfeldt, K., Maechler, P. & Wollheim, C. Molecular targets of a human HNF1 alpha mutation responsible for pancreatic beta-cell dysfunction. *EMBO Journal* **19**, 4257–4264 (2000).

62. Shih, D. *et al.* Loss of HNF-1alpha function in mice leads to abnormal expression of genes involved in pancreatic islet development and metabolism. *Diabetes* **50**, 2472–2480 (2001).

63. Dukes, I. D. *et al.* Defective Pancreaticb-Cell Glycolytic Signaling in Hepatocyte Nuclear Factor-1a-deficient Mice. *Journal of Biological Chemistry* **273**, 24457–24464 (1998).

64. Vitart, V. *et al.* SLC2A9 is a newly identified urate transporter influencing serum urate concentration, urate excretion and gout. *Nature genetics* **40**, 437–42 (2008).

65. Campbell, M. P., Royle, L., Radcliffe, C. M., Dwek, R. a & Rudd, P. M. GlycoBase and autoGU: tools for HPLC-based glycan analysis. *Bioinformatics* **24**, 1214–6 (2008).

66. Plackett, R. L. & Burman, J. P. *The design of optimum multifactorial experiments*. 305–325 (Biometrika Trust: 1947).

67. Studies, L. *Transforming Glycoscience : A Roadmap for the Future Committee on Assessing the Importance and Impact of Glycomics and Glycosciences ; Board on Chemical Sciences and Technology ; Board on Life Sciences ; Division on Earth and Life Studies ; National Resea*. (2012).

68. Aulchenko, Y. S., Ripke, S., Isaacs, A. & Van Duijn, C. M. GenABEL: an R library for genome-wide association analysis. *Bioinformatics* **23**, 1294–6 (2007).

69. Chen, W.-M. & Abecasis, G. R. Family-based association tests for genomewide association scans. *American journal of human genetics* **81**, 913–26 (2007).

70. Barrett, J. C., Fry, B., Maller, J. & Daly, M. J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263–5 (2005).

71. Davis, W. W. & Flegal, K. M. Comparing the Areas under Receiver Operating Characteristic Curves for Cluster Designsv. *Joint Statistical Meetings - Section on Survey Research Methods* 1182–1189 (2003).

72. Cardon, L. R. & Palmer, L. J. Population stratification and spurious allelic association. *Lancet* **361**, 598–604 (2003).

73. Yang, J. *et al.* Genomic inflation factors under polygenic inheritance. *European journal of human genetics : EJHG* **19**, 807–12 (2011).

74. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997–1004 (2008).

75. Higai, K., Miyazaki, N., Azuma, Y. & Matsumoto, K. Transcriptional regulation of the fucosyltransferase VI gene in hepatocellular carcinoma cells. *Glycoconjugate journal* **25**, 225–35 (2008).

76. Elliott, P. *et al.* Genetic Loci associated with C-reactive protein levels and risk of coronary heart disease. *JAMA : the journal of the American Medical Association* **302**, 37–48 (2009).

77. Yuan, X. *et al.* Population-Based Genome-wide Association Studies Reveal Six Loci Influencing Plasma Levels of Liver Enzymes. *The American Journal of Human Genetics* **83**, 520–528 (2008).

78. Kathiresan, S. *et al.* Common variants at 30 loci contribute to polygenic dyslipidemia. *Nature genetics* **41**, 56–65 (2009).

79. Reiner, A. P. *et al.* Common coding variants of the HNF1A gene are associated with multiple cardiovascular risk phenotypes in community-based samples of younger and older European-American adults: the Coronary Artery Risk Development in Young Adults Study and The Cardiovascula. *Circulation: Cardiovascular genetics* **2**, 244–54 (2009).

80. Ma, B., Simala-Grant, J. & Taylor, D. Fucosylation in prokaryotes and eukaryotes. *Glycobiology* **16**, 158R–184R (2006).

81. Boyd, M., Bressendorff, S., Møller, J., Olsen, J. & Troelsen, J. T. Mapping of HNF4alpha target genes in intestinal epithelial cells. *BMC gastroenterology* **9**, 68 (2009).

82. McDonald, T. J. *et al.* High-sensitivity CRP discriminates HNF1A-MODY from other subtypes of diabetes. *Diabetes Care* **34**, 1860–1862 (2011).

83. Bingham, C. *et al.* The Generalized Aminoaciduria Seen in Patients With Hepatocyte Nuclear Factor-1 a Mutations Is a Feature of All Patients With Diabetes and Is Associated With Glucosuria. *Diabetes* **50**, 2047–2052 (2001).

84. Stride, A. *et al.* Serum amino acids in patients with mutations in the hepatocyte nuclear factor-1 alpha gene. *Diabetic medicine* **21**, 928–30 (2004).

85. Cervin, C. *et al.* An investigation of serum concentration of apoM as a potential MODY3 marker using a novel ELISA. *Journal of internal medicine* **267**, 316–21 (2010).

86. Richter, S. *et al.* Regulation of Apolipoprotein M Gene Expression by MODY3 Gene Hepatocyte Nuclear Factor-1a. *Diabetes* **52**, 2989–2995 (2003).

87. Skupien, J. *et al.* Evaluation of Apolipoprotein M Serum Concentration as a Biomarker of HNF-1alpha MODY. *The review of diabetic studies* **4**, 231–5 (2007).

88. Owen, K. R. *et al.* Assessment of High-Sensitivity C-Reactive Protein Levels as Diagnostic Discriminator of Maturity-Onset Diabetes of the Young Due to HNF1A Mutations. *Diabetes Care* **33**, 1919–1924 (2010).

89. Thanabalasingham, G. *et al.* A large multi-centre European study validates high-sensitivity C-reactive protein (hsCRP) as a clinical biomarker for the diagnosis of diabetes subtypes Title. *Diabetologia* **54**, 2801–2810 (2011).

90. Thanabalasingham, G. *et al.* Mutations in HNF1A Result in Marked Alterations of Plasma Glycan Profile. *Diabetes* **Epub ahead**, 1–9 (2012).

91. Gornik, O. *et al.* Stability of N-glycan profiles in human plasma. *Glycobiology* **19**, 1547–53 (2009).

92. Besser, R. E. J. *et al.* Urinary C-peptide creatinine ratio is a practical outpatient tool for identifying hepatocyte nuclear factor 1-a/hepatocyte nuclear factor 4-a maturity-onset diabetes of the young from long-duration type 1 diabetes. *Diabetes care* **34**, 286–91 (2011).

93. Pal, A. *et al.* Evaluation of Serum 1,5 Anhydroglucitol Levels as a Clinical Test to Differentiate Subtypes of Diabetes. *Diabetes Care* **33**, 252–257 (2010).

94. McDonald, T. J. *et al.* Islet autoantibodies can discriminate maturity-onset diabetes of the young (MODY) from Type 1 diabetes. *Diabetic Medicine* **28**, 1028–33 (2011).

8 SUMMARY

The majority of all membrane and secreted proteins, as well as numerous cytoplasmic proteins, have one or several specific branched oligosaccharide chains (glycans) attached to their backbone. Those proteins are referred to as glycoproteins and the process of oligosaccharide attachment to a protein is called glycosylation. Glycans have numerous important structural, functional and regulatory roles in various physiological processes including protein degradation, folding and secretion, cell signaling, immune function and transcription. Glycans on the cell surface are the primary attachment site for the majority microorganisms and in fact all interactions that take place at the cell surface are modulated by glycans.

One of the modifications in the process of the glycan maturation in vertebrate, which mostly occurs in the trans-Golgi, is the fucosylation (addition of the fucose at the branched oligosaccharide), both core and antennary. Fucose can be added in an α1-6 linkage to the N-acetylglucosamine adjacent to aspargine in the core and in an α1-3 linkage as “capping” to the one or more branches (antennae).

Variations in the glycosylation of a single glycoprotein can have very profound and different consequences on its function. Absence of only one monosaccharide in the complex oligosaccharide structure can alter the function of the protein.

Glycans have only recently been introduces as potential disease biomarkers. Progress has been hindered by the lack of competent analytical methods. Alterations in glycosylation are associated with a wide range of diseases and a number of different glycan biomarkers have been identified.

Glycans can be analyzed from the body fluids or individual glycoproteins and the ability to capture and detect the diversified and heterogeneous array of glycans from the glycome is highly dependent on the availability of robust, high-resolution and sensitive separation techniques.

Generally accepted method for quantitation of reductively animated glycans is chromatography. Glycans released from the glycoprotein are labeled with fluorescent dye, which enables femtomole levels of oligosaccharide detection. Hydrophilic interaction chromatography (HILIC), a variant of the normal-phase chromatography, is the method of choice for the glycan analysis.

HNF1A-MODY is a monogenic type of diabetes, caused by mutations within 10 exons, promoters and splice sites of transcription factor hepatocyte nuclear factor (HNF)-1A that regulates gene expression in pancreatic islets, liver, kidney and intestine and is one of the key regulators of metabolic genes.

By performing the first genome-wide association analysis (GWAS) of protein glycosylation first step towards the mapping of the complex network of genes that regulate protein N-glycosylation is made. A gene encoding the transcription factor HNF1a, with previously unknown biological links to glycosylation, is shown to be strongly associated with the relative proportions of plasma N-glycans fucosylation.

Results presented in this work imply that HNF1a transcriptionally regulates the expression of genes involved in the steps of fucosylation. The present study confirms the hypothesis that the glycan profile of plasma proteins is altered substantially in those with HNF1A mutations.

The DG9-glycan index as a biomarker for HNF1A-MODY represented the best overall result although all glycan peaks with significantly lower level of antennary fucosylated structures should be further examined using larger cohorts of potential misdiagnosed patients. The DG9-glycan index is a compelling example of the potential for rapid clinical translation of a genetic discovery originating from a genome wide association analysis

9 SAŽETAK

Većina membranskih, kao i brojni citoplazmatskih proteina, sadrže jedan ili više kovalentno vezanih razgranatih lanaca ugljikohidrata (glikana). Takvi proteini nazivaju se glikoproteinima a proces vezanja različitih glikana za protein naziva se glikozilacija. Glikani imaju brojne važne strukturne, funkcionalne i regulatorne uloge u različitim fiziološkim procesima, uključujući degradaciju proteina, njihovo smatanje i sekreciju, staničnu signalizaciju, imunološki odgovor i transkripciju. Glikani na površini stanice su primarna mjesta vezanja većine mikroorganizama i zapravo sve interakcije koje se odvijaju na površini stanice modulirane su glikanima.

Jedna od modifikacija strukture u procesu sazrijevanja glikana, koja se uglavnom događa u Glogijevom aparatu, je fukozilacija (vezivanje fukoze za glikansku strukturu). Fukoza može biti vezana za N-acetilglukozamin ili preko α1-6 veze za jezgru glikana ili preko α1-3 veze za jednu ili više ogranaka (antena).

Varijacije u glikozilaciji jednog glikoproteina mogu imati vrlo značajne i različite posljedice na njegovu funkciju. Nedostatak samo jednog monosaharida u kompleksnoj strukturi jednog glikana može značajno promijeniti funkciju proteina.

Potencijal glikana kao biomarkera uočen je tek nedavno nakon razvoja novih analitičkih pristupa i metoda. Glikani se mogu analizirati u heterogenom uzorku proteina kao što je plazma ili analiza može biti usmjerena na specifičnu glikozilaciju jednog pročišćenog proteina. Oba pristupa zahtijevaju robusne, visoko protočne i osjetljive metode. Promjene glikozilacije povezane su s velikim brojem bolesti i u međuvremenu su identificirani brojni glikanski biomarkeri.

Metoda izbora za kvantitativnu analizu glikana je kromatografija. Enzimski oslobođeni glikani obilježavaju se fluorescentnom bojom što omogućava femtomolsku razinu detekcije. Najraširenija kromatografska metoda je tekućinska kromatografija hidrofilnih interakcija.

HNF1A-MODY (*maturitiy onset diabetes of the young*) je podtip dijabetesa uzrokovan mutacijama u hepatocitnom nuklearnom faktoru (HNF)-1A. To je transkripcijski faktor koji regulira ekspresiju gena u gušterači, jetri, bubrezima i probavnom traktu, a predstavlja jednog od ključnih regulatora metaboličkih gena.

Provedena je prva cijelogenomska asocijacijska studija (GWAS) glikozilacije proteina kojom se napravio prvi ključni korak prema mapiranju kompleksne mreže gena

uključenih u taj proces. Uočena je visoka povezanost gena koji kodira transkripcijski faktor HNF1A s fukozilacijom plazmatskih proteina. Ova uloga HNF1A nije od prije poznata. .

Rezultati prikazani u ovom radu pokazuju da transkripcijski faktor HNF1A regulira ekspresiju gena uključenih u proces fukozilacije i da su glikanski profili pacijenata s mutacijom u genu za HNF1A značajno promijenjeni.

Među glikanskim pikovima koji su pokazali statistički značajnu razliku između pacijenata s HNF1A-MODY tipom i ostalim tipovima dijabetesa, DG9-glikanski indeks pokazuje najbolji dijagnostički potencijal kao biomarker za HNF1A-MODY, DG9-glikanski indeks predstavlja izvrstan primjer kako se rezultati cijelogenomskih asocijacijskih studija mogu efikasno i brzo koristiti u radu s pacijentima.

10 CURRICULUM VITAE

EDUCATION

2010 –
University of Josip Juraj Strossmayer, Osijek, Molecular biosciences, Interdisciplinary postgraduate study

2000-2006
University of Zagreb, Faculty of Science, Biology Department – Molecular Biology
Graduate Study – Masters Degree Equivalent Honors

PROFESSIONAL EXPERIENCE

2009
Genos Ltd., DNA Analysis Laboratory, Glycobiology Laboratory

2007-2009
PLIVA Croatia Ltd., Research and development (Zagreb), Biotechnology, Researcher, development of follow-on Biologics

2006-2007
University of Zagreb, Faculty of Science, Department of Biology, Laboratory of Molecular Microbiology; teacher's assistant

RESEARCH AND TRAINING

2009
Water's Users Meeting – „HPLC troubleshooting", Slovenia, participant

April-December2006
Ruđer Bošković Institute, Division of Molecular Biology, Laboratory for Neurochemistry and Molecular Neurobiology (head: Dr. Branimir Jernej), Postgraduate Research

April - December 2004
Institute for Biochemistry and Molecular Biology, Charite campus Benjamin Franklin, Freie Universitaet, Berlin; trainee

2003 - 2004
University of Zagreb, Faculty of Science, Department of Botany
Undergraduate Research

HONORS

2004
Chancellor Award

2002 - 2004
Scholarship from the city of Zagreb

2001
Scholarship from the Croatian Ministry of Science and Technology

PUBLICATIONS

Thanabalasingham, G.*, Huffman, J.*, Kattla, J.*, **Novokmet, M.***, Rudan, I., Gloyn, A.L., Hayward, C., Adamczyk, B., Reynolds, R., Mužinić, A., Hassanali, N., Pučić, M., Bennett, A., Essafi, A., Polašek, O., Mughal, S., Redžić, I., Primorac, D., Zgaga, L., Kolčić, I., Hansen, T., Gasperikova, D., Tjora, E., Strachan, M., Nielsen, T., Stanik, J., Klimes, I., Pedersen, O., Njolstad, P., Wild, S.H., Gyllensten, U., Gornik, O., Wilson, J.F., Hastie, N., Campbell, H., McCarthy, M., Rudd, P., Owen, K., Lauc, G., Wright, A. Mutations in HNF1A result in marked alterations of plasma glycan profile // Diabetes, in press (2013).
*__Authors contributed equally to this work__

Lauc G, Huffman JE, Pučić M, Zgaga L, Adamczyk B, Mužinić A, **Novokmet M**, Polašek O, Gornik O, Krištić J, Keser T, Vitart V, Scheijen B, Uh HW, Molokhia M, Patrick AL, McKeigue P, Kolčić I, Lukić IK, Swann O, van Leeuwen FN, Ruhaak LR, Houwing-Duistermaat JJ, Slagboom PE, Beekman M, de Craen AJ, Deelder AM, Zeng Q, Wang W, Hastie ND, Gyllensten U, Wilson JF, Wuhrer M, Wright AF, Rudd PM, Hayward C, Aulchenko Y, Campbell H, Rudan I.Loci associated with N-glycosylation of human immunoglobulin g show pleiotropy with autoimmune diseases and haematological cancers // PLoS Genet. 2013 Jan; 9(1):e1003225. doi: 10.1371/journal.pgen.1003225

Novokmet M, Pučić M, Redžić I, Mužinić A, Gornik O. Robustness testing of the high throughput HPLC-based analysis of plasma N-glycans. // Biochim Biophys Acta. 1820(2012), 9, 1399-404.

Pučić, Maja; Mužinić, Ana; **Novokmet, Mislav**; Škledar, Marijana; Pivac, Nela; Lauc, Gordan; Gornik, Olga. Changes in plasma and IgG N-glycome during childhood and adolescence. // Glycobiology. 22 (2012) , 7; 975-982

Saldova, R.; Huffman, J.E.; Adamczyk, B.; Mužinić, Ana; Kattla, J.J.; Pučić, Maja; **Novokmet, Mislav**; Abrahams, J.L.; Hayward, C.; Rudan, Igor; Wild, S.H.; Wright, A.F.; Polašek, Ozren; Lauc, Gordan; Campbell, H.; Wilson, J.F.; Rudd, P.M. Association of medication with the human plasma N-glycome. // Journal of proteome research. 11 (2012) , 3; 1821-1831

Huffman, J.E.; Knežević, Ana; Vitart, V.; Kattla, J.; Adamczyk, B.; **Novokmet, Mislav**; Igl, W.; Pučić, Maja; Zgaga, Lina; Johannson, A.; Redžić, Irma; Gornik, Olga; Zemunik, Tatijana; Polašek, Ozren; Kolčić, Ivana; Pehlić, Marina; Koeleman, C.A.; Campbell, S.; Wild, S.H.; Hastie, N.D.; Campbell, H.; Gyllensten, U.; Wuhrer, M.; Wilson, J.F.; Hayward, C.; Rudan, Igor; Rudd, P.M.; Wright, A.F.; Lauc, Gordan. Polymorphisms in B3GAT1, SLC9A9 and MGAT5 are associated with variation within the human plasma N-glycome of 3533 European adults. // Human molecular genetics. 20 (2011) , 24; 5000-5011

Pučić, Maja; Knežević, Ana; Vidič, Jana; Adamczyk, Barbara; **Novokmet, Mislav**; Polašek, Ozren; Gornik, Olga; Šupraha Goreta, Sandra; Wormald, Mark R; Redžić, Irma; Campbell, Harry; Wright, Alan; Hastie, Nicholas D; Wilson, James F; Rudan, Igor; Wuhrer, Manfred; Rudd, Pauline M; Josić, Đuro; Lauc, Gordan. High throughput isolation and glycosylation analysis of IgG – variability and heritability of the IgG glycome in three isolated human populations. // Molecular & cellular proteomics. 10 (2011) , 10; M111.010090.

Pivac, Nela; Knežević, Ana; Gornik, Olga; Pučić, Maja; Igl, Wilmar; Peeters, Hilde; Crepel, An; Steyaert, Jean; **Novokmet, Mislav**; Redžić, Irma; Nikolac, Matea; Novković Hercigonja, Vesna; Dodig Ćurković, Katarina; Ćurković, Mario; Nedić, Gordana; Muck-Šeler, Dorotea; Borovečki, Fran; Rudan, Igor; Lauc, Gordan. Human plasma glycome in attention-deficit hyperactivity disorder and autism spectrum disorders. // Molecular & cellular proteomics. 10 (2011) , 1; 1-7

Igl, Wilmar; Polašek, Ozren; Gornik, Olga; Knežević, Ana; Pučić, Maja; **Novokmet, Mislav**; Huffman, Jennifer; Gnewuch, Carsten; Liebisch, Gerhard; Rudd, Pauline M.; Campbell, Harry; Wilson, James F.; Rudan, Igor; Gyllensten, Ulf; Schmitz, Gerd; Lauc, Gordan. Glycomics meets lipidomics—associations of N-glycans with classical lipids, glycerophospholipids, and sphingolipids in three European populations. // Molecular biosystems. 7 (2011) , 6; 1852-1862.

Knežević, Ana; Gornik, Olga; Polašek, Ozren; Pučić, Maja; Redžić, Irma; **Novokmet, Mislav**; Rudd, Pauline M; Wright, Alan F.; Campbell, Harry; Rudan, Igor; Lauc, Gordan. Effects of aging, body mass index, plasma lipid profiles, and smoking on human plasma N-glycans. // Glycobiology. 20 (2010) , 8; 959-969.

Lauc, Gordan; Essafi, Abdelkader; Huffman, Jennifer E.; Hayward, Caroline; Knežević, Ana; Kattla, Jayesh J.; Polašek, Ozren; Gornik, Olga; Vitart, Veronique; Abrahams, Jodie L.; Pučić, Maja; **Novokmet, Mislav**; Redžić, Irma; Campbell, Susan; Wild, Sarah H.; Borovečki, Fran; Wang, Wei; Kolčić, Ivana; Zgaga, Lina; Gyllensten, Ulf; Wilson, James F.; Wright, Alan F.; Hastie, Nicholas D.; Campbell, Harry; Rudd, Pauline; Rudan, Igor. Genomics meets glycomics - The first GWAS study of human N-glycome identifies HNF1α as a master regulator of plasma protein fucosylation. // Plos genetics. 6 (2010) , 12; e1001256-1-e1001256-14

Pučić, Maja; Pinto, Sofie; **Novokmet, Mislav**; Knežević, Ana; Gornik, Olga; Polašek, Ozren; Vlahoviček, Kristian; Wang, Wei; Rudd, Pauline M,; Wright, Alan F.; Campbell, Harry; Rudan, Igor; Lauc, Gordan. Common aberrations from the normal human plasma N-glycan profile. // Glycobiology. 20 (2010) , 8; 970-975.

Hranilović, Dubravka; Novak, Ruđer; Babić, Marina; **Novokmet, Mislav**; Bujas-Petković, Zorana; Jernej, Branimir. Hyperserotonemia in autism : the potential role of 5HT-related gene variants. // Collegium Antropologicum. 32 (2008) , S1; 75-80.

Mokrović, Gordana; Matošić, Ana; Hranilović, Dubravka; Štefulj, Jasminka; **Novokmet, Mislav**; Orešković, Darko; Balija, Melita; Marušić, Srđan; Čičin-Šain, Lipa. Alcohol dependence and polymorphisms of serotonin-related genes: association studies. // Collegium Antropologicum. 32 (2008) , Suppl. 1; 127-131.

Šeruga Musić, Martina; **Novokmet, Mislav**; Novak, Ruđer; Krajačić, Mladen; Škorić, Dijana. Distribution of phytoplasma diseases in the Lombardy poplar tree population of Zagreb urban area. // Acta Botanica Croatica. 67 (2008) , 2; 119-130.