Biomarker discovery in diabetes mellitus and lipid metabolism: multi-platform glyco(proteo)mic approaches

Demus, D.A.

Biomarker discovery in diabetes mellitus and lipid metabolism:

Multi-platform glyco(proteo)mic approaches

Daniel A. Demus

ISBN: 978-94-6510-142-2

The work presented in this thesis was performed at the Center for Proteomics and Metabolomics, Leiden University Medical Center, Netherlands, and at Ludger Ltd., United Kingdom.

The work was supported by the European Union Horizon 2020 Glycosylation Signatures for Precision Medicine Project, GlySign, grant number 722095.

Cover design: Daniel Demus

Biomarker discovery in diabetes mellitus and lipid metabolism: Multi-platform glyco(proteo)mic approaches

Proefschrift

ter verkrijging van

de graad van doctor aan de Universiteit Leiden,

op gezag van rector magnificus prof.dr.ir. H. Bijl,

volgens besluit van het college voor promoties

te verdedigen op dinsdag 1 oktober 2024

klokke 11:30 uur

door

Daniel Demus

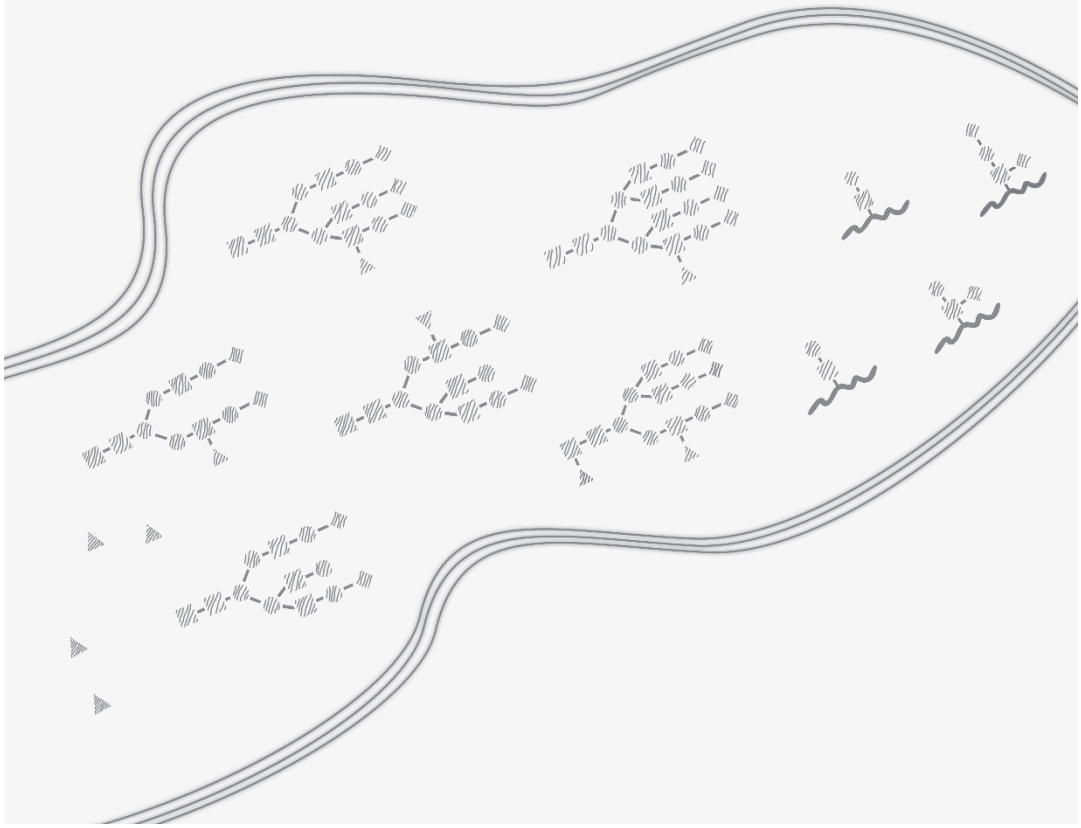
geboren te Lubaczów, Polen

in 1991

Promotor	**Prof. dr. M. Wuhrer**
Co-promotores	**Dr. D. Spencer** *Ludger Ltd., Culham Science Centre, Abingdon, the United Kingdom* **Dr. M. van Hoek** *Department of Internal Medicine, Erasmus MC University Medical Center, Rotterdam, The Netherlands*
Leden promotiecommissie	
	Prof. dr. K. Willems van Dijk **Dr. L.R. Ruhaak** **Prof. dr. C.J.J. Tack** *Radboud University, Nijmegen, The Netherlands* **Prof. dr. E.F.C. van Rossum** *Erasmus MC University Medical Center, Rotterdam, The Netherlands* **Dr. M. Baerenfaenger** *Vrije Universiteit Amsterdam, Amsterdam, The Netherlands*

Table of Contents

CHAPTER 1



1. Introduction

Glycosylation

Glycosylation is a co-/post-translational modification of proteins or lipids that occurs in the endoplasmic reticulum and Golgi apparatus of eukaryotic cells. It involves the enzymatic addition of sugar molecules (glycans) to specific sites of these molecules. Glycans are built from single monosaccharide units that combine through glycosidic bonds to form larger carbohydrates. The formation of glycans is mediated by specific enzymes – glycosyltransferases - which catalyse the synthesis of glycosidic linkages. These enzymes are classified into different families based on their amino acid sequence and the type of linkage they form between the donor and acceptor molecules. Some of the most important human glycosyltransferases belong to the families of *O*-linked glycosyltransferases (OGTs), *O*-linked *N*-acetylglucosamine (GlcNAc) and *O*-linked *N*-acetylgalactosamine (GalNAc) transferases, that catalyse the addition of a single GlcNAc and GalNAc, respectively, to the hydroxyl group of serine or threonine residues in proteins, forming *O*-linked glycosylation. Enzymes that belong to the families of *N*-linked glycosyltransferases (NGTs) form *N*-linked glycosylation by transferring an *N*-glycan via its reducing-end GlcNAc to the amide group of asparagine residues in proteins. These glycosyltransferases initiate the synthesis of two distinctive types of carbohydrate molecules: *N*-glycans and *O*-glycans[1].

N-glycans are complex, branched carbohydrate molecules that consist of a core structure of two GlcNAc and three mannose (Man) monosaccharide units. This core structure is typically further extended by the addition of various monosaccharides, including galactose, GlcNAc, *N*-acetylneuraminic acid, and fucose, resulting in a wide range of *N*-glycan structures with varying sizes and degrees of branching.

Unlike *N*-glycans, which have a conserved branched structure, *O*-glycans are highly diverse in terms of their size, composition, and branching. The biosynthesis of *O*-glycans begins with the addition of a single GalNAc residue, although other monosaccharides such as GlcNAc can also serve as initial substrates. The GalNAc

residue can then be further extended by the action of specific glycosyltransferases and glycosidases, resulting in a diverse array of core *O*-glycan structures.

The presence of *N*-glycan and *O*-glycan variants on proteins plays important roles in protein folding, secretion, stability and function. Glycoproteins on cell surface are involved in cell-cell recognition, signalling and adhesion, thus are involved in most biological processes[2]. Of note, adjusting a composition of glycan structures has found the application in biologics development to manipulate the stability and potency of therapeutic molecules[3].

Glycoproteins

Glycans are present on most human proteins and nearly all blood plasma proteins are known to be glycosylated[4]. Although there is a broad variety of glycoproteins circulating in the body, the most common classes are *N*-linked and *O*-linked glycoproteins, and glycoproteins containing multiple *N*- and *O*-glycosylation sites. The composition of glycans, which are attached to these proteins, generates proteins diversity and modulates their properties[5].

A common example of a glycoprotein whose functional characteristics are defined by *N*-glycan composition changes is a human immunoglobulin G (IgG). Afucosylated IgG is a glycoprotein that contains *N*-glycan structure in the IgG-Fc domain which lacks a fucose residue linked α1,6 on the first GlcNAc in the core. These afucosylated IgG molecules present the enhanced interaction of IgG with FcγRIII and thereby their activity. This characteristic has been broadly recognised in the biopharma industry and utilised to produce therapeutic antibodies with low core fucose content, which promotes the elevated killing activity through antibody-dependent cellular cytotoxicity mechanism of the immune system[6].

Acute phase proteins (APPs) are a major group of blood plasma proteins that undergo quantitative or qualitative changes during various pathophysiological conditions. Many of these APPs are glycosylated and it has been shown that glycosylation patterns can

be altered in pathophysiological conditions including acute inflammation, chronic inflammation (diabetes mellitus, inflammatory bowel disease), cancer (pancreatic, ovarian, lung, prostate)[7]. These glycosylation alterations correlate with disease severity in certain conditions and have been investigated as potential biomarkers. The most commonly observed APP glycosylation changes are alterations of the extent and type of fucosylation, sialylation and the number of antennae[8].

Apolipoproteins are protein components of lipoproteins - lipid-protein complexes that carry lipids in circulation, classified based on the lipid and apolipoprotein composition - that play a role in lipoprotein assembly, lipid transport, and receptor recognition[9]. Human plasma apolipoproteins, designated as A, B, (a), C, D, E, F, H, J, L, M, and O, can be classified into two categories: non-exchangeable, including ApoB-100, ApoB-48, and Apo(a), and exchangeable apolipoproteins having the ability to be transferred between different lipoprotein particle classes. Among them, apolipoproteins A, B, C, and L exhibit multiple subtypes. Glycosylation is a prevalent modification observed in nearly all human apolipoproteins and their aberrant glycosylation patterns have been linked to various diseases including metabolic syndrome[10], cancer[11], type 2 diabetes[12], autoimmune disease[13], and neurological disorders[14]. Glycosylation exerts diverse modulatory effects on apolipoprotein function, including ER-Golgi trafficking, lipoprotein assembly, stability, receptor binding and enzyme activity. For instance, the activity of lipoprotein lipase (LPL), an enzyme that degrades circulating triglycerides, is affected by the absence of *N*-glycans on apoD and the presence of sialylated *O*-glycans on apoC-III[15].

Glycosylation analysis

Studying and characterizing structures and compositions of glycans that are attached to protein or lipids is possible by applying several analytical methods. Several types of ultra-high performance liquid chromatography (UHPLC) separation techniques are used to separate and identify glycans and glycoproteins. Enzymatically or chemically released and fluorescent-labelled glycans are required for UHPLC-based analysis. The

most commonly used fluorescent tags used for glycan labelling are 2-aminobenzoic acid (2-AA), 2-aminobenzamide (2-AB) and procainamide that allow detection and relative quantification of chromatographically separated glycoforms[16, 17]. Due to the hydrophilic nature as well as size and charge variation, glycans can be separated using several molecule-stationary phase interaction mechanisms employed in UHPLC analytical columns. Hydrophilic interaction liquid chromatography (HILIC), which provides superior separation and reproducibility, has become the gold standard separation technique applied to released glycans and been applied in the majority of large-scale studies researching glycosylation changes by UHPLC methods[18, 19]. Nevertheless, reverse phase (RP)[20] and porous graphitised carbon (PGC)[21] have been proven applicable in UHPLC-based analysis of released glycans. Identification of fluorescent-labelled and chromatography separated glycans is possible by applying commercially-available glycan standards. However, UHPLC systems coupled with mass spectrometry (MS) or tandem mass spectrometry (MS/MS) detection - often supported by exoglycosidase digestions - enable the identification of rare glycan structures and overcome common hurdles of the released glycans analysis by UHPLC, such as co-eluting glycan species[22]. Similarly to the analysis of released glycans, UHPLC systems coupled with MS/MS detection can be applied to separate and identify glycoproteins. Glycoproteins can be analysed by RP, size-exclusion chromatography (SEC) and mixed-mode separation techniques.

Upon chromatographic separation, the composition of either glycan or glycoprotein samples can be analysed by electrospray ionization (ESI) MS that involves ionization of the analytes in a volatile solution. Electrospray occurs as the analyte solution is introduced into the mass spectrometer via a spray needle in the strong electric field by the application of a high voltage (2–6 kV), which causes the dispersion of the sample solution into highly charged electrospray (ES) droplets. As the solvent evaporates from the droplets in the presence of nitrogen (drying gas), multiple charged ions are formed which are transferred into the vacuum. Depending on the type of mass spectrometer used, ions with different mass-to-charge ratios (*m/z*) are then separated using a

combination of electric and magnetic fields and detected. ESI is a suitable technique for the MS analysis of glycans and glycoproteins as this “soft” ionization type does not cause significant fragmentation of these biomolecules. Identification of analytes is possible by studying their m/z values and ion fragmentation patterns, in cases when the parent ion fragmentation is applied. Different fragmentation types are available, with either collision-induced dissociation (CID) or higher-energy collisional dissociation (HCD) and electron-transfer dissociation (ETD) combinations being the most suitable for the analysis of glycans and glycoproteins[23].

Matrix-assisted laser desorption/ionization (MALDI) MS is a commonly used mass spectrometry technique for the analysis of glycans and glycoproteins. In MALDI, a matrix, which is added to a sample, absorbs the laser energy and helps to desorb and ionize the analyte molecules. MALDI is considered a “soft” ionization technique, which in most cases, allows ionization of analytes without fragmenting or decomposing them[24]. However, the labile nature of sialic acids (SA) that are present on glycans makes MALDI analysis of glycans and glycoproteins challenging. Due to the loss of sialic acids upon MALDI, several SA derivatization (stabilization) methods have been developed for the analysis of released glycans[25, 26]. Although the application of SA derivatization methods compromise some of the benefits of MALDI analysis, such as the simplicity of sample preparation, it brings the advantage of distinguishing between SA linkages through which these residues are attached to glycans. This advantage has been used in several large sample cohort studies employing MALDI analysis of released *N*-glycans and enabled evaluation of distinctive α2,3- and α2,6-linkage specific derived sialylation traits[27, 28]. In some cases the use of "cold" MALDI matrices may minimize the loss of SA residues from glycans or glycoproteins, but it does not completely prevent it [29].

Several types of plate-based assays, such as lectin microarrays and enzyme-linked lectin assays, have been developed and applied to study glycosylation. These assays typically employ specific antibodies or lectins for the capture and detection of specific glycan structures on a solid support to enable the characterization and quantification of

glycans present in biological samples[30]. A novel type of assays for the analysis of glycosylation features are enzymatic plate-based assays. These assays rely on the utilization of specific enzymes, including exoglycosidases, to catalyse reactions in microtiter plates and the subsequent quantification by detecting fluorescence signals[31].

Glycomics

Glycomics is a relatively young scientific field that is a subset of glycobiology, which aims to study complex sugars or glycans present in biological systems. The primary objective of glycomics is to gain understanding of the role that glycans play in biological processes, including molecular interactions involved in cellular communication and disease development. *N*-glycomics refers to the analysis of *N*-linked glycans, whereas *O*-glycomics specifically focuses on the study of the full repertoire of *O*-glycans, by identifying and quantifying these molecules in biological samples. In the past decade, clinical glycomics has been recognised for its contribution to understanding the biological and functional significance of glycans in various physiological and pathological processes. In the field of clinical glycomics, the analysis of a large number of samples is often required for cohort studies. Therefore, the high-throughput nature of both MALDI methods and plate-based assays makes them particularly suitable for such studies. Clinical glycomics frequently relies on the implementation of highly-automated analytical workflows supported by liquid handling platforms that employ a range of different analytical methods[32, 33]. Such workflows have been utilised to generate, as of now, blood plasma and IgG *N*-glycome datasets from large cohorts of clinical samples[28, 33]. By the identification of glycosylation aberrations in a significant number of patients, it is possible to propose new diagnostic and predictive biomarkers, and support the development of novel therapeutic approaches in various disease settings. Of note, with the advancements in analytical methods used in glycomics, particularly ultra-high resolution MS, there has been an increase in studies focusing on glycoproteomics, which involves analysing proteins that are glycosylated[34, 35].

Glycan biomarkers

Glycan biomarker refers to specific glycans or glycosylation patterns (often referred to as direct or derived traits, respectively) that are found to be associated with particular physiological or pathological conditions. Glycan biomarkers are identified through the combination of glycomics and association analysis by measuring and comparing the biomarker levels between disease and control groups of individuals. Glycan biomarkers have been shown to have the potential use in diagnosis, prognosis, and monitoring of various diseases, including cancer[36], autoimmune diseases[37], and infectious diseases[38]. It is noteworthy that *N*-glycomes of both human plasma and immunoglobulin G (IgG) remain stable over time under homeostatic conditions within an individual[39, 40]. The fact that these *N*-glycomes are highly sensitive to various pathological processes strengthens their diagnostic and prognostic potential[41, 42].

Diabetes

Diabetes mellitus is a group of chronic metabolic conditions, all sharing a common feature of heightened levels of glucose in the bloodstream. This elevation occurs due to the body's impaired capacity to produce insulin, resistance to insulin action, or both. Type 1, type 2, gestational diabetes and a group of other types of diabetes caused by genetic defects, pancreatic disease or other external diseases and influences are clinically distinct types of diabetes mellitus[43]. Type 1 diabetes refers to 5-10% of diabetes cases characterised by the autoimmune destruction of beta cells within the pancreas, which results in the complete absence of insulin production. Type 2 diabetes accounts for 90-95% of all diabetes cases and develops in cases of an abnormal increased resistance to the action of insulin, and when the body fails to produce enough insulin to overcome the resistance. If left untreated, diabetes can cause long term complications such as kidney damage, nerve damage, cardiovascular disease, and vision loss[44].

The group of diabetes types, which are distinct from type 1, type 2 and gestational diabetes, includes monogenic diabetes syndromes. Maturity-onset diabetes of the

young (MODY), a rarer type of diabetes consisting of less than 5% diabetes cases, is one of the main forms of monogenic diabetes caused by a single gene mutation. There are 14 known subtypes of MODY, each caused by mutations in different genes that leads to distinct clinical features[45].

Disease complications are defined as medical conditions that arise during a disease or are a result of unfavorable evolution of an existing disease, a health condition, or a medical treatment. Long term complications of untreated or inadequately regulated diabetes, classified as microvascular or macrovascular, pose a significant burden on public health systems. Microvascular complications affect nervous system (neuropathy), renal system (nephropathy) and eye health (retinopathy) whereas macrovascular complications are the cause of stroke, cardiovascular disease and peripheral vascular disease[46].

Glycan biomarkers in diabetes

Several studies suggest that protein *N*-glycosylation may play a role in the development of type 2 diabetes[47, 48]. *N*-glycosylation changes have been observed in various diseases such as type 1 diabetes, type 2 diabetes and HNF1A-MODY[49]. These *N*-glycosylation changes are considered as potential biomarkers, which can differentiate between different types of diabetes and predict the risk of developing diabetes in the future.

A large cohort study identified that individuals at a higher risk of developing type 2 diabetes can be identified based on variations in glycosylation. An increased complexity of *N*-glycan structures, specifically highly branched, galactosylated, and sialylated structures in total plasma proteins, was found to be associated with an elevated risk of developing type 2 diabetes[47]. Another study has found that a proinflammatory state, which is characterised by a decrease in galactosylation and sialylation of IgG, is observed in individuals with type 2 diabetes. Moreover, an increase in pro-inflammatory IgG fucosylated structures with bisecting GlcNac was observed in these individuals[48, 50].

Several studies have shown that by measuring antennary fucosylation (α1-3,4-linked fucose residues) of *N*-glycans, it is possible to discriminate HNF1A-MODY and other types of diabetes[51, 52]. The *N*-glycan antennary fucosylation levels are significantly decreased in individuals with HNF1A-MODY as the consequence of the presence of genetic variants in the *HNF1A* gene[53]. Recently, fucosylated alpha-1-acid glycoprotein (AGP) has been evaluated for its diagnostic potential in HNF1A-MODY[54]. The best performing AGP variant was found to provide a very good discriminatory potential with an AUC of 0.94, which is in line with the results obtained in the previous studies that considered released *N*-glycans and absolute α1-3,4 fucosylation levels of total plasma proteins[51, 55].

Glycan biomarkers in complications of diabetes

There have been several reports on the associations between either total plasma *N*-glycome or IgG *N*-glycome and type 2 diabetes, however, there is a limited number of studies on the involvement of *N*-glycosylation in type 2 diabetes complications. In a recent study on two large Dutch cohorts using mass spectrometry to assess total plasma *N*-glycome, several glycosylation features, including fucosylation, galactosylation and sialylation, were found to be associated with prevalent and incident complications of type 2 diabetes. High levels of *N*-glycan bisection was strongly associated with prevalent cardiovascular disease, while the increase in 2,6-sialylation on tri-antennary glycans was observed in cases of prevalent nephropathy[28].

Dyslipidemia, a cardiovascular risk factor prevalent in approximately 50% of individuals with type 2 diabetes, is characterised by a preponderance of small, dense, low-density lipoprotein particles, elevated triglyceride and low high-density lipoprotein cholesterol levels[56]. Alterations in apolipoprotein glycosylation have been linked to various diseases associated with dyslipidemia, including diabetes[57]. Apolipoprotein C-III (ApoC-III) is one of the main four forms of ApoC found in chylomicrons, VLDL, and HDL, and ApoC-III *O*-glycosylated variants have been found to be associated with fasting plasma triglyceride levels[58]. The complete molecular mechanisms underlying these

associations have not been fully elucidated. However, it is noteworthy that sialylated apoC-III glycoforms play a role in the differential clearance of triglyceride-rich lipoproteins[59]. These findings suggest that apoC-III has the potential to serve as a biomarker and be targeted for the treatment of lipoprotein metabolism disorders.

Scope of the thesis

The aim of the research presented in this thesis is development and optimization of analytical methods to study glycosylation changes as potential stratification biomarkers in large clinical sample cohorts of patients with two types of diabetes and diabetes-related complications. The analytical workflows proposed in this research were designed to address and overcome challenges in areas of high-throughput sample preparation, the analysis of these samples, subsequent data processing and statistical analysis. The developed methods are applicable for the analysis of released *N*-glycans, *O*-glycosylated proteins and absolute fucosylation levels of proteins derived from blood plasma.

Since the first study in 2013 which proposed that blood plasma protein *N*-glycan antennary fucosylation is altered in patients with a monogenic type of diabetes, HNF1A-MODY, there have been several studies evaluating the potential of full plasma protein antennary fucosylated *N*-glycans as stratification biomarkers to differentiate patients with HNF1A-MODY from other diabetes cases. In **chapter 2**, a new LC-MS/MS method was developed to assess *N*-glycan antennary fucosylation levels in a large number of patients and perform a first inter-laboratory evaluation of this glycan biomarker by comparing the biomarker's diagnostic performance and the consistency of *N*-glycan antennary fucosylation level measurements that were performed in two independent research centres.

The application of LC methods is poorly recognised in public diagnostic centres, which sets a major bottleneck in translating glycan biomarkers into clinical practice. In **chapter 3**, an enzymatic plate-based assay, which employs high-throughput sample preparation using a liquid-handling platform and simplified data processing, was developed to assess absolute α1-3,4 fucosylation levels of blood plasma proteins in a large cohort of patients with HNF1A-MODY and type 2 diabetes. The diagnostic performance of α1-3,4 fucosylation levels was evaluated against the diagnostic performance of antennary fucosylated *N*-glycans obtained in the previous studies applying LC methods.

Elevated triglyceride levels are observed in diabetic dyslipidemia. Altered profiles of apoC-III *O*-glycosylated sialylated variants have been linked to increased plasma triglyceride levels based on results from several studies applying various analytical methods. Small sample cohorts that were applied in the current studies might be a significant concern in consideration of the validity of their research discoveries. In **chapter 4**, a previously established highly-automated ultra-high resolution MALDI-FTICR MS method was adjusted to facilitate the analysis of apoC-III *O*-glycosylation profiles in large sample cohorts. The workflow that was applied to collect, process and curate MALDI MS-derived data involves the use of in-house developed high-throughput processing software, MassyTools. The association analysis was performed for apoC-III *O*-glycosylation profiles and a panel of lipid biomarkers.

The method for the analysis of apoC-III *O*-glycosylation, which was optimised and validated in **chapter 4**, was applied for a well characterised large cohort of patients with type 2 diabetes. ApoC-III *O*-glycosylation profiles were obtained for these patients, including a control group of patients without diabetes, and undergone genome-wide association studies (GWASs) to identify associations between genetic variants and apoC-III glycoform levels in **chapter 5**. Further, apo-CIII *O*-glycosylation profiles derived from groups of patients with type 2 diabetes, which were categorised based on the presence of micro- and macrovascular complications, were subjected to the association analysis to uncover associations between apo-CIII *O*-glycoform levels and either the prevalence or the incidence of these complications in **chapter 6**. The research findings were strengthened by performing the meta-analysis of genetic variants linked to apoC-III glycosylation with lipid biomarkers and micro-/macro-vascular complications of type 2 diabetes in **chapter 5 and chapter 6**, respectively.

In the final chapter, **chapter 7**, research studies presented in this thesis are discussed and the research findings critically evaluated in the context of their significance for clinical use. Aspects that pose a challenge in translating glycan biomarkers into clinical practice are highlighted. Moreover, the potential of analytical methods and statistical

approaches employing large omics data, which are proposed in this thesis, is evaluated for its use in the biopharma industry.

CHAPTER 2

2. Interlaboratory evaluation of plasma *N*-glycan antennary fucosylation as a clinical biomarker for HNF1A-MODY using liquid chromatography methods

Daniel Demus[1,2], Bas C. Jansen[1], Richard A. Gardner[1], Paulina A. Urbanowicz[1], Haiyang Wu[3], Tamara Štambuk[4,5], Agata Juszczak[6], Edita Pape Medvidović[7], Nathalie Juge[3], Olga Gornik[5], Katharine R. Owen[6,8], Daniel I. R. Spencer[1]

[1] Ludger Ltd, Culham Science Centre, Abingdon, Oxfordshire, England, United Kingdom

[2] Center for Proteomics and Metabolomics, Leiden University Medical Center, Leiden, The Netherlands

[3] Quadram Institute Bioscience, Norwich Research Park, United Kingdom

[4] Genos Glycoscience Research Laboratory, Zagreb, Croatia

[5] Faculty of Pharmacy and Biochemistry, University of Zagreb, Zagreb, Croatia

[6] Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Oxford, Oxfordshire, England, United Kingdom

[7] Vuk Vrhovac University Clinic for Diabetes, Endocrinology and Metabolic Diseases, Merkur University Hospital, Zagreb University School of Medicine, Zagreb, Croatia School of Medicine, Zagreb, Croatia

[8] Oxford NIHR Biomedical Research Centre, Oxford Hospitals NHS Foundation Trust, Oxford, Oxfordshire, England, United Kingdom

Reprinted and adapted: Demus, D., Jansen, B.C., Gardner, R.A. et al. Interlaboratory evaluation of plasma N-glycan antennary fucosylation as a clinical biomarker for HNF1A-MODY using liquid chromatography methods. Glycoconj J 38, 375–386 (2021). https://doi.org/10.1007/s10719-021-09992-w

ABSTRACT

Antennary fucosylation alterations in plasma glycoproteins have been previously proposed and tested as a biomarker for differentiation of maturity onset diabetes of the young (MODY) patients carrying a functional mutation in the *HNF1A* gene. Here, we developed a novel LC-based workflow to analyse blood plasma *N*-glycan fucosylation in 320 diabetes cases with clinical features matching those at risk of HNF1A-MODY. Fucosylation levels measured in two independent research centres by using similar LC-based methods were correlated to evaluate the interlaboratory performance of the biomarker. The interlaboratory study showed good correlation between fucosylation levels measured for the 320 cases in the two centres with the correlation coefficient (r) of up to 0.88 for a single trait A3FG3S2. The improved chromatographic separation allowed the identification of six single glycan traits and a derived antennary fucosylation trait that were able to differentiate individuals carrying pathogenic mutations from benign or no *HNF1A* mutation cases, as determined by the area under the curve (AUC) of up to 0.94. The excellent (r = 0.88) interlaboratory performance of the glycan biomarker for HNF1A-MODY further supports the development of a clinically relevant diagnostic test measuring antennary fucosylation levels.

INTRODUCTION

Glycosylation is a co-/post-translational modification which affects protein conformation and interactions. Glycans can be covalently attached to proteins through either *N*- or *O*-glycosidic linkages. These oligosaccharide chains strongly influence protein-protein interactions and are involved in protein folding, sub-cellular targeting, and trafficking. *N*-glycans are linked to the amide group of asparagine (Asn), whereas *O*-glycans are attached to the hydroxyl group of serine (Ser) or threonine (Thr) in the polypeptide backbone[2].

Aberrant protein glycosylation has been associated with many pathological conditions in humans. Some chronic inflammatory[60], autoimmune[61] and infectious diseases[38], such as inflammatory bowel disease[62], rheumatoid arthritis[63] and HIV infection[64] were associated with changes in *N*- and/or *O*-glycosylation profiles of plasma proteins. Glycan signatures have been extensively studied in the field of cancer research, with altered sialylation, galactosylation[65] and fucosylation[66] identified in various cancer types. These studies showed the potential of using glycans as biomarkers to screen for certain pathological conditions, monitor patients undergoing treatment or predict occurrence or reoccurrence of a disease even before any physiological symptoms can be detected[67-72].

HNF1A-MODY (MODY3) is a type of maturity onset diabetes of the young (MODY) for which a glycan biomarker has been proposed for patient identification and stratification. HNF1A-MODY is the most common form of autosomal dominant monogenic diabetes in adults, accounting for around 50% of all MODY cases in the UK[45]. It is caused by mutations in the *HNF1A* gene which encodes for the HNF-1α protein, a transcription factor regulating expression of several genes that are important for pancreatic β-cell development and function[73]. Different rare variants (classified as damaging, likely-damaging and benign) in the *HNF1A* gene cause various severity of the disease. Damaging (loss-of-function, pathogenic) mutations in the *HNF1A* cause deregulation of β-cell genes and these mutations have been shown to result in

generation of abnormal β-like cells in vitro[74]. The disease is characterised by progressive deterioration of β-cell function, but insulin sensitivity remains normal or slightly increased[75, 76]. Clinically, individuals with HNF1A-MODY are at increased risk of vascular complications due to hyperglycaemia[77].

Although glycomics studies have demonstrated the ability to systematically differentiate HNF1A-MODY patients from other diabetes forms[51, 53], there has not yet been widespread adoption of this approach in clinical environments and currently, testing for MODY depends on the awareness and interest of individual clinicians. A MODY probability calculator based on clinical features has been developed to help clinicians identify those with the highest risk of MODY[78]. The presence of *HNF1A* mutations requires genetic testing but the diverse range of mutations is not completely predictive of changes in the protein function[79, 80]. In addition, due to the cost and unavailability of genetic testing in many countries, many patients remain misdiagnosed or unidentified. Importantly, misdiagnosis often results in patients being referred to an inappropriate course of treatment[81].

Genome-wide association studies (GWAS) have shown that C-reactive protein (CRP) expression and *N*-glycan antennary fucosylation levels are significantly altered in patients harbouring mutations in the *HNF1A* gene[82]. Changes in antennary fucosylation are a result of altered activity of specific fucosyltransferases catalysing the α-1,3 or α-1,4 addition of fucose in the antennary *N*-acetylglucosamines of *N*-glycans. Previous studies that have determined the levels of *N*-glycan fucosylation in blood plasma using liquid chromatography (LC) methods showed the excellent performance of the glycan biomarker in differentiating HNF1A-MODY patients in large patient cohorts[51, 53].

Due to the evolution of liquid separation techniques since the first HNF1A-MODY study[53], it is now possible to perform a more exhaustive evaluation of potential glycan biomarkers. Furthermore, *N*-glycan antennary fucosylation as a biomarker for HNF1A-MODY has not been evaluated in the context of its robustness and repeatability of

measurements between laboratories. Assessment of interlaboratory performance of the biomarker will provide further valuable insights into its clinical application.

Here, we evaluated the interlaboratory performance of plasma protein antennary fucosylation as a biomarker for HNF1A-MODY based on a set of 320 patients with non-autoimmune diabetes from the cohort previously analysed[51]. We analysed glycans present in the blood plasma by using a newly developed workflow employing a LC method with fluorescence (FD) and mass spectrometric (MS/MS) detection (LC-FD-MS/MS) supported by exoglycosidase digestions. Subsequently, we compared fucosylation levels for 320 individuals that were measured in the current study and analysed previously as a part of the study by Juszczak et al.[51]using LC-based methods in two research centres. In addition to the individual *N*-glycan structures, a derived antennary fucosylation trait was calculated and tested for its diagnostic potential.

MATERIALS AND METHODS

Sample cohort

The research presented here was performed using blood plasma samples obtained from study participants that were recruited via the Young Diabetes in Oxford (YDX) study in the UK (n=90) and the Croatian National Diabetes Registry (CroDiab) in Croatia (n=230). Subjects older than 18 years who were diagnosed with diabetes before age of 45 years were eligible to take part in the study. Biochemical inclusion criteria were fasting C-peptide ≥ 0.2 nmol/L, which indicates insulin production, and negative GAD antibodies (GADA: the commonest antibody found in type 1 diabetes). Informed consent was obtained for all participants. Clinical characteristics of patients included in the study are summarised in **Table 1**.

Plasma samples were obtained from patients' blood. All cases underwent Sanger sequencing for HNF1A-MODY as part of the MODY-glycan study[51]. Plasma samples were grouped based on *HNF1A* mutation type, which was allocated using a systematic and functional assessment of rare *HNF1A* alleles. These included three mutation groups: (likely) damaging, variants of unknown significance (VUS), (likely) benign and a group of patients without *HNF1A* mutation (D).

Table 1. Clinical characteristics of patients included in the study.

	(Likely) damaging allele n = 18 cases	VUS n = 5 cases	(Likely) benign n = 8 cases	No rare *HNF1A* allele variant n = 289 cases
Sex, male:female [%]	28:72	40:60	50:50	52:48
Age at recruitment [years]	39.7 (17.4)	54.2 (14.2)	48.1 (11.2)	45.6 (9.7)
Age at diagnosis [years]	25.7 (8.6)	37.3 (9.8)	36.4 (5.6)	35.3 (6.5)
Diabetes duration [years]	13.1 (11.6)	17.3 (10.6)	9.1 (6.6)	10.3 (7.7)
BMI [kg/m2]	25.93 (4.75)	28.34 (6.19)	34.61 (5.02)	29.95 (5.76)
hsCRP [mg/L]	0.80 (1.32)	4.01 (6.30)	5.21 (5.75)	3.79 (6.97)
HbA1c [%]	7.7 (1.7)	6.7 (1.3)	8.8 (2.1)	7.6 (1.8)
C-peptide [nmol/L]	0.42 (0.21)	0.60 (0.57)	0.81 (0.62)	0.75 (0.36)
Total cholesterol [mmol/L]	4.75 (1.09)	4.74 (1.21)	5.07 (0.88)	4.79 (1.36)
HDL [mmol/L]	1.37 (0.32)	1.28 (0.39)	1.11 (0.23)	1.17 (0.36)
Triglycerides [mmol/L]	1.24 (0.51)	1.12 (0.31)	1.98 (0.96)	2.15 (1.98)

Mean values are presented for each group (standards deviation (SD) in parentheses).

N-glycan release, fluorescent labelling and purification

Enzymatic release and fluorescent labelling of *N*-glycans from human blood plasma samples were performed using a highly automated analytical workflow for high-throughput (HTP) glycomics supported by a Hamilton STARlet liquid handling robot, as described previously[83, 84] with minor modifications. Here, we used 4 µL of human blood plasma samples and a procainamide labelling kit with sodium cyanoborohydride as the reductant (LT-KPROC-96, Ludger) for fluorescent labelling of released *N*-glycans.

Exoglycosidase digestion

Procainamide-labelled plasma *N*-glycan samples (8 µL) were treated with either 1 µM E1_10125 α-L-fucosidase from Ruminoccocus gnavus strain E1 (E1_10125) with specificity towards α1-3,4>2 fucose[85] or 1 µL bovine kidney α-L-fucosidase (BKF; Sigma; specific towards α1-6>2,3,4 fucose)[86]. The enzymatic reaction was carried out in sodium phosphate buffer (50 mM, pH 6.0) at 37°C for 18 hours. Digested *N*-glycans were purified using LT-KPROC-96 plate (Ludger), eluted in 200 µL of water, evaporated to dryness and reconstituted in 100 µL of water. During the procedure, an enzymatically untreated aliquot of procainamide-labelled plasma *N*-glycans was used as a process control.

HILIC-LC-FD-MS/MS analysis of PROC-labelled glycans

Procainamide-labelled *N*-glycan samples were analysed by LC-FD-MS/MS with electrospray ionization (ESI). Prior to analysis, samples were freshly prepared in a 96-deep well collection plate (LP-COLLPLATE-2ML-96; Ludger) on the Hamilton STARlet liquid handling robot as a mixture of water and acetonitrile in a 1:3 ratio. To verify optimal LC-MS signals, a procainamide-labelled plasma *N*-glycan standard from a readily available source was used as a system suitability standard prior to each analysis. The readily available standard had been prepared from pooled human plasma (Sigma Aldrich) according to a standard protocol (described in ***N*-glycan release, fluorescent labelling and purification**), aliquoted, stored at -21°C before use.

Each sample (25 µL) was injected onto a HALO 2 Penta-HILIC 150 × 2.1 mm column with 2.0 µm stationary phase particle size (AMT91812-705, Advanced Materials Technology) at 40 °C on a Dionex Ultimate 3000 UHPLC instrument with a fluorescence detector (excitation and emission wavelengths of 310 and 370 nm, respectively), coupled in-line to a Bruker Amazon Speed ETD mass spectrometer. Data acquisition was controlled by HyStar version 3.2 (Bruker). Solvent A was 200 mM ammonium formate buffer (pH 4.4) (LS-N-BUFFX40, Ludger), solvent B was acetonitrile and solvent C was pure water.

A 70-min run was used with a multi-step gradient consisting of 7-44% solvent A, 72-56% solvent B, 21-0% solvent C over 62 min at a flow rate of 0.4 ml/min, followed by 44-50% A, 56-0% B, 0-50% C in 3.5 min at a flow rate of 0.3 mL/min, returning to 7% A, 72% B and 21% C over 4 min at a flow rate of 0.45 mL/min. The mass spectrometer was used in enhanced resolution mode, positive ion setting, at range of 400-1700 m/z. Other settings were as follows: nebulizer pressure 14.5 psi, capillary voltage 4500 V, nitrogen flow 10 L/min, ion charge control (ICC) target 200000, maximum accumulation time 50.00 ms, singly charged ions excluded.

Data processing and statistical analysis

Bruker Compass DataAnalysis version 4.1 and Bruker ProteinScape version 4.0 software (GlycomeDB database) were used to analyse mass spectrometry (MS) data. Glycan structures were identified based on combination of LC-FD-MS/MS data supported by exoglycosidase digestion data, the UHPLC column separation characteristics and data from the literature[87]. GlycoWorkBench version 2 was used for searching most plausible glycan structures based on accurate mass. Fluorescence traces were exported as an open text format using Chromeleon version 7.2 (Thermo Fisher). UHPLC data processing and quantification was performed using HappyTools version 0.0.2 build 180521a[88]. A feature list, containing peak retention times and widths, was generated using the automated peak detection option of HappyTools. Subsequently, the feature list was manually curated after visual inspection of the overlaid chromatograms. The main features were selected and used for calibration of all the chromatograms, while the entire feature list was used for HTP quantification of all the detectable features

(minimum intensity = 0.001; Sigma edge method; Sigma value = 2.0) in the chromatograms. Analyte and chromatogram QC parameters were also determined using HappyTools. These parameters were then used to support manual curation of the results, evaluating the Gaussian peak quality (GPQ ≥ 0.8) and signal-to-noise (SN ≥ 9) ratio for each LC feature.

Statistical analysis and visualization were performed using Microsoft Excel and R platform version 1.1.463. Bonferroni correction was applied to test for sex and age effect. All tests were significant after Bonferroni correction (number of tests n = 183; 6 potential markers with sex, age and, sex*age correction; p cut-off = 2.78x10-2). For Receiver Operating Characteristic (ROC) curve analysis, (likely) damaging, (likely) benign and no *HNF1A* mutation groups were tested against each other. Single glycan traits were tested for the best classification performance. Area under the curve (AUC), optimal sensitivity, specificity and cutoff values were generated using cutpointr R package. A derived trait, which averages antennary fucosylation feature across individual glycan structures, was tested separately. Data outliers were identified with the 1.5xIQR rule[89]. Outliers were detected for each glycan trait per patient group before ROC analysis.

Data collected as a part of the previous study by Juszczak et al.[51] using the same sample cohort was used to evaluate the interlaboratory performance of the glycan biomarker by Spearman correlation method. In the previous study, the relative abundance of glycans in each glycan peak was expressed as a percentage of the total integrated area. Here, fucosylation levels were expressed in fucosylation indexes that were calculated for each glycan trait (**Table 2**). The best performing glycan traits, corresponding to the same glycan structure (GP30 – A3FG3S2, GP36 – A3FG3S3, GP38 – FA3FG3S3) in both studies, were subjected to the correlation analysis.

RESULTS

Development and validation of a novel workflow for identification and characterization of antennary fucosylation of human plasma proteins

To determine the antennary fucosylation levels of 320 subjects with diabetes, we developed a workflow encompassing highly reproducible ultra-high-performance liquid chromatography (UHPLC) analysis and high-throughput (HTP) systematic data processing. This approach allows a highly robotised, simultaneous analysis of 96 samples made up of 93 clinical samples and a pooled human plasma standard in triplicate as a control.

Our LC-FD-ESI-MS/MS-based approach included the use of a HILIC-type UHPLC column that separates glycans depending on their size and degree of sialylation coupled with exoglycosidase digestions for glycan structure identification (**Supplementary Figure S1**). The use of a column packed with core–shell silica particles, which provides lower backpressure, was implemented to achieve a desired glycan separation without excessively long run times[90]. Definitive identification of fucosylated *N*-glycan structures was achieved using a combination of FD-MS/MS detection and exoglycosidase digestions. Analysis of the fragmentation mass spectra (MS) was performed with a focus on the following ions: m/z 512.19 for the GlcNAc-Gal-Fuc fragment and m/z 587.33 for the PROC-GlcNAc-Fuc fragment, which indicate antennary and core fucosylation of *N*-glycans, respectively. Since fucose migration may lead to misleading readouts of MS/MS data[91, 92], exoglycosidase digestions were used to detect and distinguish the presence of antennary and core fucosylation. The penta-HILIC column provided separation of various glycan isomers, specifically sialic acid linkage isomers, revealing two tetra-sialylated antennary-fucosylated structures (A4FG4S4). We conclude that the two A4FG4S4 glycan structures are sialic acid linkage isomers, however determination of sialic acid linkages (α-2,3 or α-2,6) is not possible as no sialic acid derivatization approach was applied in this study. Forty-nine glycan peaks were assigned (**Supplementary Figure S2 and Supplementary Table S1**) and 6 peaks were identified as antennary fucosylated structures: p29, p35, p42, p44, p48, p49,

corresponding to A2FG2S2, A3FG3S2, A3FG3S3, FA3FG3S3 and two isomers of A4FG4S4 respectively.

To evaluate technical variations within the analytical approach, pooled human plasma was used as an internal control in triplicate in order to monitor consistency of sample preparation and data processing steps. The triplicates were also used to evaluate repeatability of the glycan analysis (**Supplementary Table S2**). The most abundant peak was p28 with a coefficient of variation (CV) of 1%. Peaks with relative intensities (RI) above 1% (p25, p28, p29, p32, p34, p38, p41, p42, p46 and p47) gave the average CVs of 7% and 5% for the inter- and intra-plate studies, respectively.

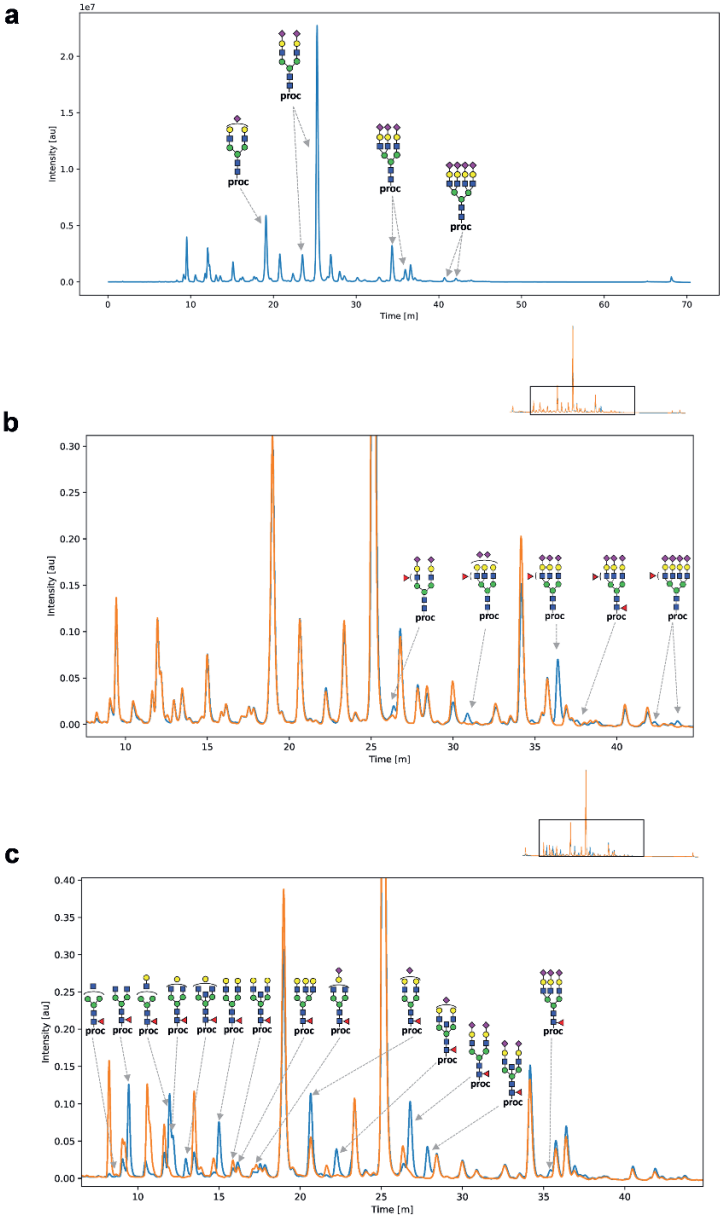


Figure S1. A representative fluorescence chromatogram of released procainamide-labelled plasma *N*-glycans with main glycan peaks assigned (a) and fluorescence chromatograms obtained after exoglycosidase treatment with E1_10125 (b) and BKF (c). Displayed glycan structures indicate that the column groups glycans depending on their degree of sialylation. Chromatograms of enzymatically treated samples (orange) were overlaid with control samples (blue) and normalised to the highest peak. Zoomed areas show the glycan structure peak shifts. A multi-tool approach was applied to assign glycan structures to LC peaks. Structure assignment was based on both retention time and mass detection including manual examination

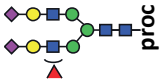
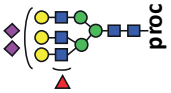
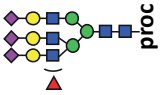
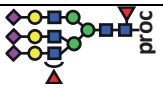
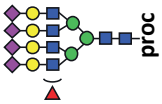
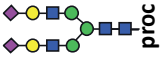
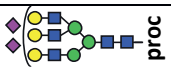
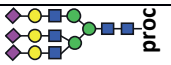
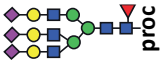
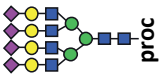
of the MS/MS data. Fluorescence chromatograms obtained after treatment with two fucosidases with different substrate specificities were analysed for peak shifts that resulted in changes of peak relative intensity in comparison to controls (b-c). For graphical representations of glycan structures: blue square (*N*-acetylglucosamine), green circle (mannose), yellow circle (galactose), purple diamond (*N*-acetylneuraminic acid), red triangle (fucose). Antennary fucosylated glycan peaks and their non-fucosylated counterpart peaks were used for calculations of traits that were subsequently applied to measure fucosylation levels in blood plasma samples. The glycan peaks used for analyses and calculated traits are listed and summarised in Table 2. Glycosylation traits were calculated based on relative peak areas normalized to the peaks used for statistical analysis (**Supplementary Table S2**).

Table S2. Technical variation of plasma N-glycan peaks used for statistical analysis with average relative intensities (RI), standard deviations (SD) and coefficient of variation (CV) values calculated based on standard plasma triplicates. The average relative intensities (RI), standard deviations (SD) and coefficient of variation (CV) were calculated based on pooled plasma triplicates for peaks used in statistical analysis within and between LC-MS runs.

Peak	***Plate 1***	*RI*	*SD*	*CV (%)*	***Plate 2***	*RI*	*SD*	*CV (%)*	***Plate 3***	*RI*	*SD*	*CV (%)*
p25		*0.0668*	*0.001*	*1%*		*0.0713*	*0.0027*	*4%*		*0.0652*	*0.0022*	*3%*
p28		*0.6474*	*0.0097*	*1%*		*0.6625*	*0.0094*	*1%*		*0.6514*	*0.0094*	*1%*
p29		*0.0175*	*0.001*	*6%*		*0.0182*	*0.0018*	*10%*		*0.0157*	*0.0004*	*2%*
p32		*0.0257*	*0.0019*	*7%*		*0.0269*	*0.0036*	*13%*		*0.0276*	*0.0019*	*7%*
p34		*0.0181*	*0.0017*	*9%*		*0.0195*	*0.0049*	*25%*		*0.0187*	*0.0015*	*8%*
p35		*0.0085*	*0.0004*	*5%*		*0.009*	*0.0018*	*20%*		*0.0093*	*0.0007*	*8%*
p38		*0.1018*	*0.0028*	*3%*		*0.0929*	*0.0033*	*4%*		*0.0989*	*0.0025*	*3%*
p40		*0.0072*	*0.0007*	*10%*		*0.0062*	*0.0004*	*6%*		*0.0062*	*0.0001*	*1%*
p41		*0.0313*	*0.0014*	*4%*		*0.0278*	*0.0008*	*3%*		*0.0302*	*0.0014*	*5%*
p42		*0.0422*	*0.002*	*5%*		*0.0378*	*0.0007*	*2%*		*0.0433*	*0.0025*	*6%*
p44		*0.0027*	*0*	*1%*		*0.0025*	*0*	*1%*		*0.0028*	*0.0001*	*3%*
p46		*0.0122*	*0.0008*	*7%*		*0.0102*	*0.0003*	*3%*		*0.012*	*0.0009*	*7%*
p47		*0.0121*	*0.0011*	*9%*		*0.0097*	*0.0003*	*3%*		*0.0117*	*0.0009*	*8%*
p48		*0.0027*	*0.0002*	*8%*		*0.0022*	*0.0003*	*11%*		*0.0025*	*0.0003*	*11%*
p49		*0.0039*	*0.0003*	*9%*		*0.003*	*0.0001*	*4%*		*0.0045*	*0.0006*	*13%*

	Plate 4	*RI*	*SD*	*CV (%)*	***All plates***	*RI*	*SD*	*CV (%)*
p25		*0.0651*	*0.0006*	*1%*		*0.0671*	*0.0029*	*4%*
p28		*0.6405*	*0.0063*	*1%*		*0.6504*	*0.0092*	*1%*
p29		*0.0158*	*0.0001*	*1%*		*0.0168*	*0.0012*	*7%*
p32		*0.0222*	*0.0002*	*1%*		*0.0256*	*0.0024*	*9%*
p34		*0.0132*	*0.0011*	*8%*		*0.0174*	*0.0028*	*16%*
p35		*0.0069*	*0.0003*	*4%*		*0.0084*	*0.0011*	*13%*
p38		*0.0958*	*0.0027*	*3%*		*0.0974*	*0.0038*	*4%*
p40		*0.0063*	*0.0006*	*10%*		*0.0065*	*0.0005*	*7%*
p41		*0.0301*	*0.0008*	*3%*		*0.0299*	*0.0015*	*5%*
p42		*0.0413*	*0.0015*	*4%*		*0.0411*	*0.0024*	*6%*
p44		*0.0028*	*0.0003*	*12%*		*0.0027*	*0.0001*	*5%*
p46		*0.0118*	*0.0003*	*3%*		*0.0115*	*0.0009*	*8%*
p47		*0.012*	*0.0004*	*3%*		*0.0114*	*0.0011*	*10%*
p48		*0.0028*	*0.0004*	*14%*		*0.0026*	*0.0003*	*10%*
p49		*0.0048*	*0.0002*	*3%*		*0.004*	*0.0008*	*19%*

Table 2. Construction of direct and derived antennary fucosylation traits.

Antennary fucosylated glycan peak	Glycan notation	Proposed glycan structure*
p29	A2FG2S2	
p35	A3FG3S2	
p42	A3FG3S3	
p44	FA3FG3S3	
p48 p49	A4FG4S4	
Non-antennary fucosylated glycan peak	**Glycan notation**	**Proposed glycan structure**
p25 p28	A2G2S2	
p32 p34	A3G3S2	
p38 p41	A3G3S3	
p40	FA3G3S3	
p46 p47	A4G4S4	

Single traits	Trait calculation
A2FG2S2	p29 / p25+p28+p29
A3FG3S2	p35 / p32+p34+p35
A3FG3S3	p42 / p38+p41+p42
FA3FG3S3	p44 / p40+p44
A4FG4S4_I	p48 / p46+p48
A4FG4S4_II	P49 / p47+p49

Derived trait	Trait calculation
Antennary fucosylation trait	$\frac{p29 + p35 + p42 + p44 + p48 + p49}{p25 + p28 + p29 + p32 + p34 + p35 + p38 + p40 + p41 + p42 + p44 + p46 + p47 + p48 + p49}$

*the graphical representations of glycan structures used the Symbol Nomenclature for Glycans (SNFG): blue square (*N*-acetylglucosamine), green circle (mannose), yellow circle (galactose), purple diamond (*N*-acetylneuraminic acid), red triangle (fucose)

Interlaboratory evaluation of the glycan biomarker for classification of HNF1A-MODY

The workflow developed and described above was applied to plasma samples collected as part of a previous cohort from patients in the UK and Croatia, described elsewhere[51]. Based on the presence and assessed disease-causality of *HNF1A* mutations, participants were divided in four groups: (likely) damaging (n = 18), (likely) benign (n = 8) and VUS (n = 5), and a group of no *HNF1A* mutation diabetes cases (n = 289). Antennary fucosylated glycan levels were measured and subsequently correlated for the same set of cases (n = 320) using similar LC-based workflows in two independent laboratories, one in the UK and one in Croatia.

We compared three directly measured, antennary fucosylated glycans – A3FG3S2 (chromatographic peak GP30 in Juszczak et al., p35 herein), A3FG3S3 (GP36, p42) and FA3FG3S3 (GP38, p44) , which showed the best discriminative performance for HNF1A-MODY in the study by Juszczak et al[51]. The analysis showed a significant correlation for the antennary fucosylated glycan levels expressed as either the relative abundance (percentage of a certain *N*-glycan structure in the total plasma *N*-glycome)[51] or as antennary fucosylation indexes (current study). The Spearman's correlation coefficient (r) ranged between 0.69 and 0.88 (**Figure 1**). The best results were obtained for antennary fucosylated structures A3FG3S2 and A3FG3S3 with r scores of 0.88 and 0.86, respectively. The core and antennary fucosylated glycan structure (FA3FG3S3) trait gave the r score of 0.69.

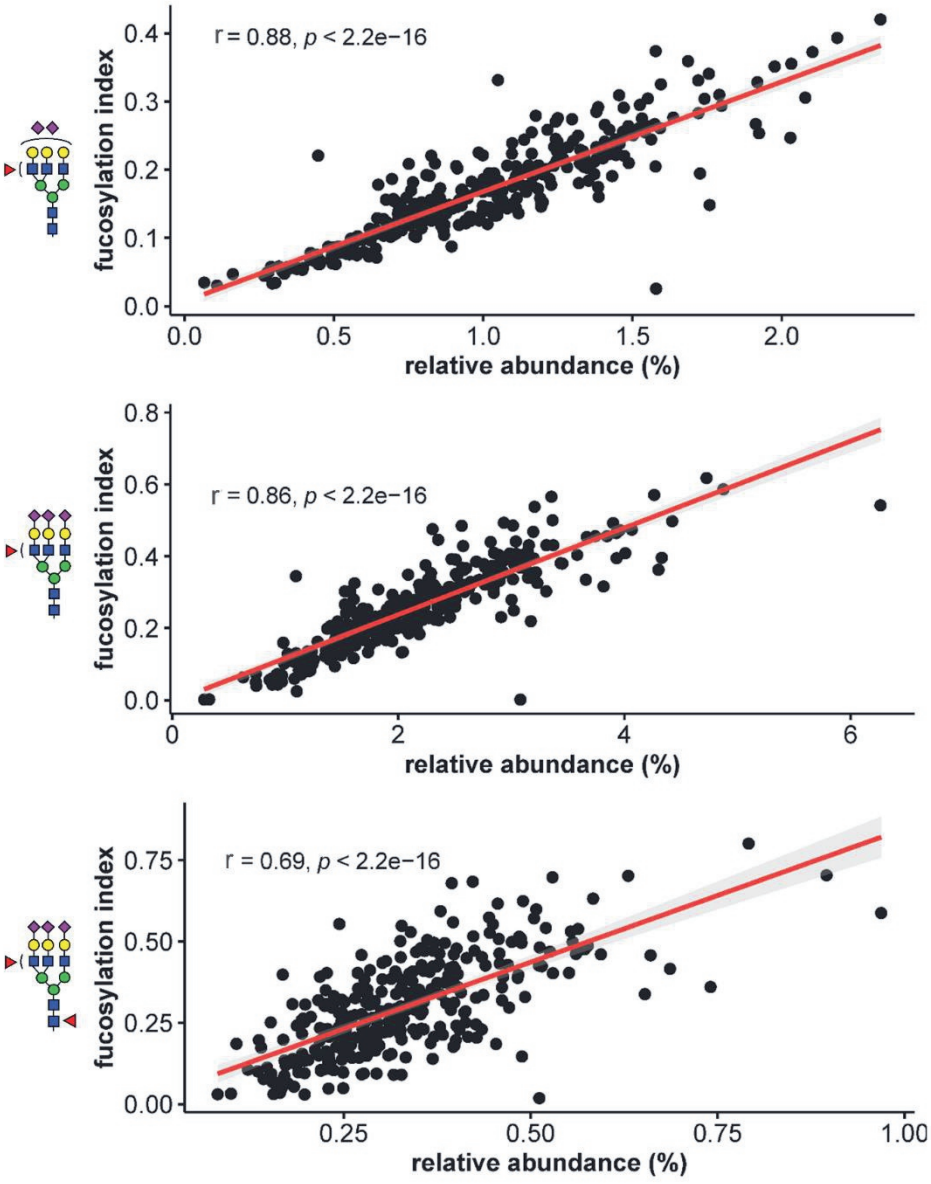


Figure 1. Correlation analysis illustrating the performance of three single glycan traits used as differentiating biomarkers for HNF1A-MODY in two independent laboratories. Antennary fucosylation levels measured as the relative abundance of antennary fucosylated glycans[51] or as fucosylation indexes (current study) were compared for 320 individuals with diabetes. The performance of each glycan trait is described by the Spearman's correlation coefficient (r).

Identification of novel glycosylation traits as biomarkers of HNF1A-MODY

Glycosylation trait analysis was carried out to identify single fucosylated glycan structures that could discriminate between a pathogenic group of (likely) damaging (n = 18) mutation cases and non-pathogenic groups of either benign (n = 8) or no *HNF1A* mutation (n = 289) cases.

We evaluated differences in antennary fucosylation levels between *HNF1A* mutation groups measured by applying single traits (A3FG3S2, A3FG3S3, FA3FG3S3, A4F4G4_I, A4G4S4_II) and the derived trait (**Figure 2**). A3FG3S2, A3FG3S3, FA3FG3S3, A4F4G4_I, A4G4S4_II and the derived antennary fucosylation trait allow for differentiation of the pathogenic mutation group from the nonpathogenic mutation groups, by showing significant differences ($p \leq 0.05$) in fucosylation levels between these groups. Individuals with (likely) damaging mutations in the *HNF1A* gene present significantly decreased antennary fucosylation levels in respect to benign and no mutation groups. No significant difference was observed between (likely) benign and no *HNF1A* mutation groups for all traits.

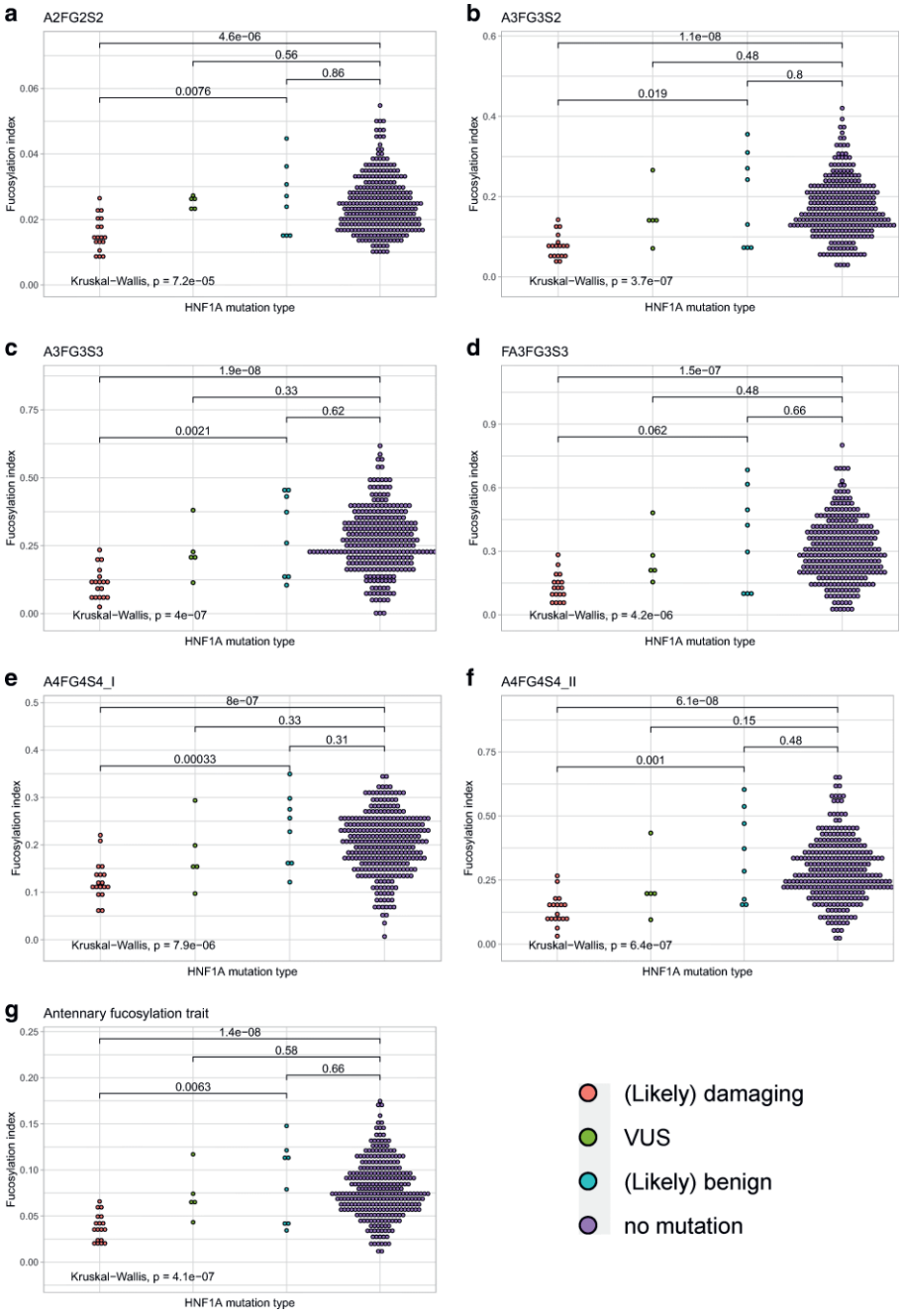


Figure 2. Dot plots presenting antennary fucosylation levels measured using each glycosylation trait for groups of individuals without *HNF1A* mutation and with different *HNF1A* mutation types. The lines and numbers above the box plots, indicate the p-value when comparing two categories using The Wilcoxon-Mann-Whitney test. The analysis with $p \leq 0.05$ is considered statistically significant.

Receiver Operating Characteristic (ROC) curves were used to assess the diagnostic accuracy of glycan traits for discrimination of *HNF1A* mutation types. As a result of data curation, out of the 320 samples tested, 24 samples (including 3 pathogenic cases) for single traits and 4 samples for the derived trait were classified as outliers prior to ROC analysis. ROC analysis showed that all tested traits provided good discriminative power between pathogenic and no *HNF1A* mutation cases, as determined by the measure of area under the curve (AUC) found to be in a range of 0.84 to 0.94 (**Supplementary Table S3**). A3FG3S2 and A3FG3S3 performed best within the single traits providing AUCs of 0.94 with 93% sensitivity, 87% specificity at the cutoff of 0.101 and 100% sensitivity and 83% specificity at the cutoff of 0.163, respectively (**Figure 3a - 3b**). The derived trait performed well in differentiating pathogenic cases from no *HNF1A* mutation group with an AUC of 0.90, 83% sensitivity and 78% specificity at the cutoff of 0.057 (**Figure 3c**). By applying these three markers, one case within the VUS group (n = 5) would have been classified as pathogenic. Two novel single traits based on isomers of A4FG4S4 glycan gave AUCs of 0.90 and 0.93 for A4FG4S4_I and A4FG4S4_II, respectively, when distinguishing between pathogenic and no *HNF1A* mutation groups. When testing (likely) damaging cases against (likely) benign group, the glycan traits gave AUCs in a range of 0.76 to 0.96 with A4FG4S4_I, providing the best discriminative power within the single traits (**Supplementary Table S3**).

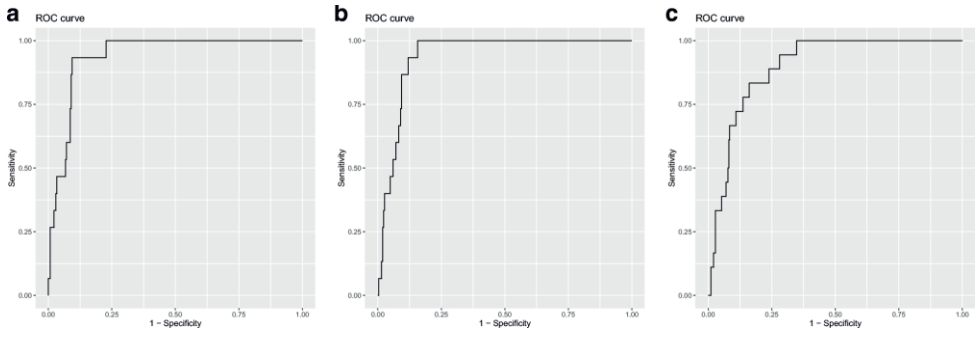


Figure 3. ROC curves illustrating the performance of single glycan traits (a and b) and the derived antennary fucosylation trait (c) in differentiating patients with (likely) damaging *HNF1A* mutation from a group of patients without *HNF1A* mutation. The AUC values are displayed for the best performing single glycan traits and the derived trait.

Table S3. Performance of glycan traits in differentiating subjects with different *HNF1A* mutation status

Groups tested	Glycan trait	AUC	Cutoff	Sensitivity	Specificity
(Likely) damaging *HNF1A* vs. no mutation	A2FG2S2	0.84	0.018	0.73	0.72
	A3FG3S2	0.94	0.101	0.93	0.87
	A3FG3S3	0.94	0.163	1.00	0.83
	FA3FG3S3	0.91	0.200	1.00	0.76
	A4FG4S4_I	0.90	0.158	0.93	0.76
	A4FG4S4_II	0.93	0.188	1.00	0.81
	Derived trait	0.90	0.057	0.83	0.78
(Likely) benign *HNF1A* vs. no mutation	A2FG2S2	0.54	0.033	0.86	0.25
	A3FG3S2	0.55	0.251	0.91	0.38
	A3FG3S3	0.58	0.381	0.90	0.38
	FA3FG3S3	0.57	0.481	0.94	0.38
	A4FG4S4_I	0.63	0.264	0.89	0.38
	A4FG4S4_II	0.60	0.421	0.93	0.38
	Derived trait	0.55	0.107	0.82	0.50
(Likely) damaging vs. (likely) benign *HNF1A*	A2FG2S2	0.85	0.023	1.00	0.63
	A3FG3S2	0.84	0.104	0.93	0.63
	A3FG3S3	0.92	0.168	1.00	0.63
	FA3FG3S3	0.76	0.224	1.00	0.63
	A4FG4S4_I	0.96	0.165	1.00	0.75
	A4FG4S4_II	0.94	0.196	1.00	0.63
	Derived trait	0.83	0.069	1.00	0.63

More than half of all antennary fucosylated *N*-glycans in human blood plasma derive from α-1-acid glycoprotein (AAG), an acute phase protein[93, 94]. Correlation between CRP, AAG, and other inflammatory proteins has been previously reported[95], therefore CRP levels were checked within (likely) damaging *HNF1A* (n = 15) and no *HNF1A* mutation (n = 267) cases included in the study, and (likely) damaging *HNF1A* (n = 3) and no *HNF1A* mutation (n = 21) cases classified as data outliers, excluded from the ROC analysis. Increased CRP levels, regarding the mean value calculated within (likely) damaging cases included in the ROC analysis, were observed for the outliers, however, the difference was not considered statistically significant (**Supplementary Table S4**). Furthermore, increased rates of α1,3-fucosylation of tri-antennary and tetra-antennary

glycans detected in sera of patients with chronic inflammation have been reported[96]. In support of these findings we observed significant positive correlations for the glycan traits based on tri- and tetra-antennary structures (A3FG3S2, A3FG3S3, FA3FG3S3, A4FG4S4_I, A4FG4S4_II) as well as the derived fucosylation trait and CRP levels within the cohort (**Supplementary Table S5**).

Table S4. CRP levels and antennary fucosylation indexes calculated for damaging *HNF1A* and no *HNF1A* mutation cases included in the study (n = 15 and 267, respectively) and damaging *HNF1A* and no *HNF1A* mutation cases classified as data outliers (n = 3 and 21, respectively) excluded from the ROC analysis.

	CRP level (mg/L)			**Antennary fucosylation index**		
	Study cases	Outliers	**p-value**	Study cases	Outliers	**p-value**
Damaging *HNF1A* mutation group	0.72 (1.42)	1.22 (0.46)	**0.075**	0.033 (0.011)	0.062 (0.004)	**0.0025**
No *HNF1A* mutation group	3.80 (7.16)	3.73 (3.90)	**0.364**	0.076 (0.026)	0.126 (0.042)	**6.3e-09**

Values are presented as: mean (SD). The Wilcoxon-Mann-Whitney test was applied to compare the means; $p \leq 0.05$ is considered statistically significant.

Table S5. The correlation analysis for glycan traits and CRP marker (n = 320).

Glycan trait	**Spearman's correlation coefficient (r)**	**p-value**
A2FG2S2	-0.034	0.54
A3FG3S2	0.14	0.011
A3FG3S3	0.15	0.006
FA3FG3S3	0.12	0.032
A4FG4S4_I	0.15	0.0079
A4FG4S4_II	0.18	0.0012
Derived antennary fucosylation trait	0.18	0.0015

$p \leq 0.05$ is considered statistically significant

DISCUSSION

Individuals with HNF1A-MODY carry variants in the *HNF1A* gene, encoding for a transcription factor HNF1α, which has been found to be a master regulator of plasma protein fucosylation[82]. Expression of genes encoding for enzymes involved in L-fucose biosynthesis and several fucosyltransferases (FUTs) is under the regulation of HNF1α. In the presence of loss-of-function variants in the *HNF1A* gene, repressing activity of HNF1α towards FUT8 decreases. This leads to upregulation of core fucosylation in an α-1,6 linkage to the first GlcNAc residue in *N*-glycans and downregulation of antennary fucosylation in α-1,3 and α-1,4 linkage positions catalysed by FUT3, FUT5 and FUT6[97]. Based on these findings, the decrease in α-1,3 and α-1,4 antennary fucosylation of *N*-glycans from blood plasma protein has been proposed as a biomarker to identify individuals with HNF1A-MODY[53].

In this study, we analysed *N*-glycan fucosylation of plasma glycoproteins in 320 individuals with diabetes and clinical features matching those at risk of HNF1A-MODY, using a newly developed LC-based workflow. The new workflow provides an excellent chromatographic separation of antennary fucosylated structures and is enhanced by the use of novel exoglycosidases. Unlike other commercially available antennary fucosidases, the novel E1_10125 enzyme is characterised by activity towards α-1,3/4 linkage fucosylated substrates presenting a terminal sialic acid modification[85]. By applying E1_10125 and the commercially available α-1,6 linkage specific fucosidase (BKF, Sigma), we were able to fully distinguish between antennary and core fucosylation of *N*-glycan structures. Retaining terminal sialic acid residues enabled the analytical method used to maintain the improved chromatographic separation power and variations in specific isoforms. Using this analytical approach, most antennary fucosylated structures found on human blood plasma proteins[87] were tested for their clinical potential as HNF1A-MODY biomarkers.

The interlaboratory performance of the glycan-based biomarker for HNF1A-MODY was determined by correlating antennary fucosylation levels measured using single glycan

traits for the same set of samples by using similar LC methods in two independent research centres. Different UHPLC columns, sample processing and glycan separation conditions were applied in both centres. 2-aminobenzamide-labelled and procainamide-labelled glycans were analysed by Juszczak et al. and in the current study, respectively[51]. Despite technical and methodological differences in the analytical workflows, the evaluation study proved the robustness and reproducibility of antennary fucosylated *N*-glycans as biomarkers for HNF1A-MODY, as shown by a strong correlation for the three glycan traits. The traits based on antennary fucosylated glycan structures (A3FG3S2 and A3FG3S3) gave very consistent readouts of fucosylation levels within 320 samples for the two laboratories with the correlation coefficient (r) ranging from 0.86 to 0.88. The trait based on the glycan structure containing both core and antennary fucosylation (FA3FG3S3) performed the worst in the interlaboratory study. Previously, the opposite effect that mutations in the *HNF1A* gene have on core and antennary fucosylation has been reported[82]. Therefore, taking into account the methodological differences such as trait calculation in both studies, we conclude that core fucosylation and the presence of an additional antennary fucosylated glycan structure (A4FG4S3) under the peak GP38 might have had an impact on the worser performance of FA3FG3S3.

Furthermore, the new analytical workflow allowed us to identify six single antennary fucosylated glycan traits that were tested for their clinical potential to discriminate between patients with pathogenic and non-pathogenic mutations in the *HNF1A* gene, or without them. The best performing single glycan traits were found to be A3FG3S2 and A3FG3S3, both giving an AUC of 0.94, for classifying pathogenic mutation cases against a group of cases without the *HNF1A* mutation. The same single glycan traits were also the best performing ones in the study by Juszczak et al.[51], achieving an AUC of 0.90 for both A3FG3S2 and A3F2G3S3. This minor difference in classification performance assessed by ROC curve analysis is due to different study design, since the previous study considered a larger group of subjects with non-autoimmune diabetes

and no *HNF1A* mutations (n=960), while herein only a subset of the group was analysed (n=289).

Thanabalasingham et al. previously proposed a derived trait (Dg-9 index), obtained by averaging antennary fucosylated tri-antennary glycans against all tri-antennary glycans, which was evaluated for the discrimination of HNF1A-MODY from groups of well-defined patients with a diagnosis based on molecular, biochemical and clinical investigation[53]. Here, the improved chromatographic separation allowed us to calculate and subsequently test the derived trait, which reflects overall changes in antennary fucosylation levels of blood plasma proteins better than a single trait. The derived antennary fucosylation trait performed well in differentiating pathogenic *HNF1A* mutation cases from the control group with an AUC of 0.90. Importantly, all single and derived antennary fucosylation traits provided a good discriminative power between pathogenic and benign mutation cases, as shown by AUCs ranging between 0.76 and 0.96. By applying these glycosylation traits, we were able to classify cases within the VUS group, in which one would have been categorised as carrying a damaging mutation type in the *HNF1A* gene, which was identified by significantly decreased fucosylation levels compared to the non-pathogenic and no mutation groups.

Significant overexpression of FUT6 has been reported in patients with non-alcoholic steatohepatitis (NASH), which is a progressive form of non-alcoholic fatty liver disease (NAFLD). Further, increased levels of fucosylated glycoprotein α-1 antitrypsin have been observed with pathological changes of NAFLD, such as fibrosis, pathological inflammation, and ballooning[98]. Due to the unavailability of specific markers of liver function in the clinical data of the collected sample cohort we were not able to determine the liver condition of individuals with diabetes enrolled in this study and a possible influence of liver dysfunction on antennary fucosylation levels. Although we have investigated a relationship between antennary fucosylation and CRP, which is also a predictor of NAFLD[99], a more in-depth understanding of inflammatory events that

have a potential influence on the biomarker's performance is needed. We acknowledge this aspect as a limitation of the current study, which should be investigated further.

In conclusion, glycan biomarkers have great potential in being applied to distinguish between healthy and diseased individuals. A large number of diseases including cancer and diabetes have been associated with changes in glycosylation profiles, which could be obtained through non- or minimally invasive analyses of body fluids and tissue samples[49, 100]. However, the expertise and high-end instrumentation that is required to detect these often subtle glycosylation changes are major obstacles in translation of novel glycan-based markers into clinical application. Moreover, the majority of diseases are the consequence of complex changes, for which combined diagnostic biomarkers are required[95]. Whereas, HNF1A-MODY is a disease for which a biomarker based solely on a glycosylation feature is sufficient to classify patients, as shown in the current and previous studies[51, 53].

Here, we have demonstrated the performance of a glycan-based biomarker for HNF1A-MODY across laboratories. This study confirms antennary fucosylation of *N*-glycans as a promising tool for patient stratification by enabling discrimination of cases with pathogenic mutations in the *HNF1A* gene compared to those with benign or variants of unknown significance. It is worth mentioning that the performance of antennary fucosylation as a biomarker for HNF1A-MODY could be influenced by inflammation events, as shown by the relationship between increased antennary fucosylation and CRP levels within (likely) damaging cases classified as outliers, therefore the significance of this observation should be further explored. Together, this is the first study which proves excellent robustness and reproducibility of the glycan biomarker for HNF1A-MODY tested on the same set of clinical samples across laboratories. The data presented here supports the development of a simpler, high-throughput assay for determining antennary fucosylation levels, which will facilitate future translation of this glycan biomarker into wider clinical practice.

Acknowledgments

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 722095. The authors gratefully acknowledge the support of the Biotechnology and Biological Sciences Research Council (BBSRC) with contribution from the Innovate UK Biocatalyst grant Glycoenzymes for Bioindustries (BB/M029042/) and the BBSRC Institute Strategic Programme Grant Gut Microbes and Health BB/R012490/1 and its constituent project BBS/E/F/000PR10356 (Theme 3, Modulation of the gut microbes to promote health throughout life). The research was supported by the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre (BRC). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. AJ was a Diabetes UK George Alberti fellow during the research.

The CroDiab study was approved by Ethics Committee of Clinical Hospital Merkur and University of Zagreb, Faculty of Pharmacy and Biochemistry, Croatia. The study was conducted in accordance with the Declaration of Helsinki.

The YDX study was approved by the Oxfordshire Local Research Ethics Committee, and all subjects gave written informed consent.

CHAPTER 3

3. Development of an exoglycosidase plate-based assay for detecting α1-3,4 fucosylation biomarker in individuals with HNF1A-MODY

Daniel Demus[1,2], Paulina A. Urbanowicz[1], Richard A. Gardner[1], Haiyang Wu[3], Agata Juszczak[4], Tamara Štambuk[5,6], Edita Pape Medvidović[7], Katharine R. Owen[4,8], Olga Gornik[6], Nathalie Juge[3], Daniel I. R. Spencer[1]

[1] Ludger Ltd, Culham Science Centre, Abingdon, Oxfordshire, England, United Kingdom

[2] Center for Proteomics and Metabolomics, Leiden University Medical Center, Leiden, The Netherlands

[3] Quadram Institute Bioscience, Norwich Research Park, United Kingdom

[4] Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Oxford, Oxfordshire, England, United Kingdom

[5] Genos Glycoscience Research Laboratory, Zagreb, Croatia

[6] Faculty of Pharmacy and Biochemistry, University of Zagreb, Zagreb, Croatia

[7] Vuk Vrhovac University Clinic for Diabetes, Endocrinology and Metabolic Diseases, Merkur University Hospital, Zagreb University School of Medicine, Zagreb, Croatia School of Medicine, Zagreb, Croatia

[8] Oxford NIHR Biomedical Research Centre, Oxford Hospitals NHS Foundation Trust, Oxford, Oxfordshire, England, United Kingdom

Reprinted and adapted: Daniel Demus, Paulina A Urbanowicz, Richard A Gardner, Haiyang Wu, Agata Juszczak, Tamara Štambuk, Edita Pape Medvidović, Katharine R Owen, Olga Gornik, Nathalie Juge, Daniel I R Spencer, Development of an exoglycosidase plate-based assay for detecting α1-3,4 fucosylation biomarker in individuals with HNF1A-MODY, Glycobiology, Volume 32, Issue 3, March 2022, Pages 230-238, https://doi.org/10.1093/glycob/cwab107

ABSTRACT

Maturity-onset diabetes of the young due to hepatocyte nuclear factor-1 alpha variants (HNF1A-MODY) causes monogenic diabetes. Individuals carrying damaging variants in *HNF1A* show decreased levels of α1-3,4 fucosylation, as demonstrated on antennary fucosylation of blood plasma *N*-glycans. The excellent diagnostic performance of this glycan biomarker in blood plasma *N*-glycans of individuals with HNF1A-MODY has been demonstrated using liquid chromatography methods. Here, we have developed a high-throughput exoglycosidase plate-based assay to measure α1-3,4 fucosylation levels in blood plasma samples. The assay has been optimised and its validity tested using 1000 clinical samples from a cohort of individuals with young-adult onset diabetes including cases with HNF1A-MODY. The α1-3,4 fucosylation levels in blood plasma showed a good differentiating power in identifying cases with damaging *HNF1A* variants, as demonstrated by ROC curve analysis with AUC values of 0.87 and 0.95. This study supports future development of a simple diagnostic test to measure this glycan biomarker for application in a clinical setting.

INTRODUCTION

Glycosylation is a co-/post-translational modification of proteins which is important for protein structure, stability and function, and influences most biological processes including cell signalling. There are two types of glycosylation: *N*-glycosylation where glycans are attached to a protein via a glycosidic bond to asparagine and *O*-glycosylation where the glycosidic bond links glycans to either serine or threonine[101]. Advances in analytical technologies have helped to further understand the role of glycosylation in human health and disease. Alterations in glycosylation patterns occur in certain pathological conditions such as cancers[36], autoimmune diseases and in response to lifestyle changes[37, 102]. Glycosylation features such as sialylation, fucosylation and galactosylation have been explored in different diseases as potential biomarkers[36] for early diagnosis[103] or for monitoring the effectiveness of treatments[68, 104]. The measurement of glycosylation features may also be applied to estimate and track the biological age of healthy individuals[105].

Maturity onset diabetes of the young due to damaging alleles in hepatocyte nuclear factor-1 alpha (HNF1A-MODY) is a rare type of diabetes caused by an autosomal dominant mutation in the single gene, *HNF1A*, which is involved in regulating β-cell development and insulin secretion[74]. HNF1A-MODY is characterised by defects in insulin secretion and onset of hyperglycemia in the 2nd-4th decade of life[81]. A correct diagnosis is a key factor for optimal disease management, as early stages of HNF1A-MODY can be effectively controlled with orally administered sulfonylureas[106]. Diagnosis of HNF1A-MODY is challenging as it often depends on the awareness of physicians and the availability of relatively expensive genetic testing for confirming the presence of pathogenic variants in the *HNF1A* gene. Moreover, HNF1A-MODY shares clinical features with other types of diabetes which is estimated to lead to misdiagnosis in around 80% of cases in the UK[79]. A MODY probability calculator has been developed to help physicians detect those at higher risk of MODY[78]. Moreover, due to the autosomal dominant inheritance pattern of HNF1A-MODY, a correct diagnosis helps identify other family members affected by this disorder[81].

Certain variants in the *HNF1A* gene result in altered glycosylation patterns associated with HNF1A-MODY. Genome-wide association studies (GWAS) show that damaging loss-of-function variants in the *HNF1A* gene lead to downregulation of fucosyltransferases that perform the α1-3 and α1-4 fucosylation of glycans[82]. Further studies proposed that decreased α1-3 and α1-4 fucosylation of *N*-glycans (antennary fucosylation) from blood plasma proteins can be used as a differentiating biomarker for HNF1A-MODY[53]. The use of antennary fucosylation as a biomarker of HNF1A-MODY was previously demonstrated[51, 53] and we recently showed the excellent inter-laboratory performance of the *N*-glycan biomarker using liquid chromatography (LC) methods[52]. Nevertheless, the applicability of LC-based methods in clinical practice is limited due to their low robustness and equipment-related costs, therefore development of a simpler analytical approach to test for this well-studied biomarker is warranted.

Here, we have developed an exoglycosidase plate-based assay which enables high-throughput measurements of levels of fucose residues that are attached via α1-3 and α1-4 linkages to glycoconjugates present in blood plasma. The assay has been optimised and its performance validated on 1000 clinical blood plasma samples from individuals with young-adult onset diabetes, including groups of individuals with different variants in the *HNF1A* gene. The differentiating power of α1-3,4 fucosylation levels is comparatively discussed with the previously published results testing antennary fucosylated *N*-glycans as a biomarker for HNF1A-MODY using LC methods.

MATERIALS AND METHODS

Material

The blood samples were obtained from the HNF1A-MODY cohort[51]. Briefly, participants of this study, were recruited from two European centres via the Young Diabetes in Oxford (YDX) study in the UK (n=499) and the Croatian National Diabetes Registry (CroDiab) in Croatia (n=501). The study inclusion criteria were the age of 18 years or older and diabetes diagnosis before the age of 45 years. Biochemical inclusion criteria were: fasting C-peptide ≥ 0.2 nmol/L, which indicates endogenous insulin production and negative GAD antibodies (GADA) to exclude type 1 diabetes patients. Informed consent was obtained from all participants. Sequencing of *HNF1A* and systematic and functional assessment of rare *HNF1A* alleles, performed as part of the previous study[51], allowed to divide the participants into four groups with (likely) damaging *HNF1A* variants , (likely) benign *HNF1A* variants, a group of cases with variants of unknown significance (VUS) and a group without variants in the *HNF1A* gene.

α1-3,4 fucosylation level measurements

Patients' blood plasma samples and pooled blood plasma standard (FRNCP0125, VisuCon) were applied in the exoglycosidase plate-based assay. The pooled plasma standard was applied in the assay optimization experiments and used as a process control in the patients cohort study. Blood plasma samples were thawed, vortexed and centrifuged briefly at 600 rpm for 30 s prior to the exoglycosidase plate-based assay. Duplicate 10 µL aliquots of each sample were transferred manually into a 96-well PCR plate (4ti-0960, 4titude) and diluted 5-fold in 50 mM citrate buffer, pH 6. The plate was then sealed with a pierce foil seal (4ti-0531, 4titude), incubated at 100°C for 10 min to enhance denaturation of proteins, then cooled down at 4°C and centrifuged briefly at 600 rpm for 30 s.

The pierce foil seal was removed before the plate was placed on a Hamilton STARlet liquid handling robot where the next steps were performed using a semi-automated program. The blood plasma samples were treated with an exoglycosidase for fucose release or left untreated (no exoglycosidase). Briefly, 11 µL α1-3,4 specific fucosidase

(E1_10125)[85] was added at 3 µM final enzyme concentration in 250 mM citrate buffer, pH 6. The dilution buffer (11 µL) was added to the untreated samples. The plate was then removed from the robot, sealed again, mixed on a plate shaker for 1 min and centrifuged at 800 rpm for 30 s. Following incubation at 37°C overnight (16 ± 1 h), the plate was cooled down at 4°C and briefly centrifuged at 600 rpm for 30 s, the seal removed and the plate was placed back in the robot where 54 µL of ultrapure water was added to each well. The plate was sealed, the samples were mixed by vortexing for 1 min and then centrifuged at 1400 rpm for 40 min.

The seal was then removed and the plate placed into the robot which transferred 54.4 µL of supernatant into a 384-well microplate (4ti-0234, 4titude). An L-fucose assay kit containing reaction buffer, NADP+, fucose dehydrogenase and fucose standard (K-FUCOSE, Megazyme), diaphorase (D5540, Sigma) and resazurin (199303, Sigma Aldrich, UK) were used in the following steps. A standard curve was prepared in 4 replicates in the 384-well microplate with L-fucose concentration ranging from 0, 0.86, 1.69, 2.52, 3.35, 4.17 and 5.00 ng/µL. A reagent mix (containing 10.9 µL of reaction buffer, pH 9.5, 5.4 µL of 1 mM resazurin solution and 2.7 µL of NADP+) and an enzymatic reagent mix (containing 1.1 µL of fucose dehydrogenase and 5.4 µL of 10 U/mL diaphorase solution) were added to each sample. The plate was then removed from the robot, sealed and incubated for 3 h in the dark at 21°C. The plate was then centrifuged at 600 rpm for 30 s and placed in a plate reader (Enspire 2300, Perkin Elmer Enspire, USA) for fluorescence measurements at 24°C with an excitation wavelength (λex) = 571 nm and an emission wavelength (λem) = 586 nm. Three consecutive measurements were taken with excitation illumination from above the plate at a height of 10 mm. Fluorescence signals from the three measurements were averaged.

Data processing and statistical analysis

Data processing, visualization and statistical analysis were performed using Microsoft Excel and R platform version 1.1.463. Receiver Operating Characteristic (ROC) curve analysis was applied to estimate diagnostic performance, testing three *HNF1A* variant groups: (likely) damaging, (likely) benign and diabetes cases without *HNF1A* mutation.

ROC curves, area under the curve (AUC), optimal sensitivity, specificity and cutoff values were generated using "cutpointr" R package. Data outliers were identified with the 1.5xIQR rule and detected per mutation group[89]. Statistical significance of differences in α1-3,4 fucosylation levels between mutation groups was determined using the Mann-Whitney U test for pairwise comparison and Kruskal-Wallis test for global comparison ($p < 0.05$). The Spearman's correlation method was used to test α1-3,4 fucosylation levels against a panel of inflammatory and metabolic markers available within the cohort. Body mass index (BMI) and C-reactive protein (CRP) level values were available in the clinical data for 866 and 916 patients, respectively. The Spearman's correlation method was used to evaluate α1-3,4 fucosylation levels of glycoconjugates in blood plasma and antennary fucosylation levels of plasma *N*-glycans measured as indexes as a part of the previous HNF1A-MODY study[52]. A generalised linear model adjusted for either age or sex and high-sensitivity CRP (hsCRP) levels as confounding variables was used to investigate associations between α1-3,4 fucosylation levels (dependent variable) and sex or age.

Chromeleon version 7.2 (Thermo Fisher) was used to export fluorescence traces as an open text format. UHPLC data processing, quantitation and graphical overlays of chromatograms were performed using HappyTools version 0.0.2 build 1800521a[88]. Microsoft Excel was used to export fluorescence readout generated by the micro-plate reader EnSpire (Perkin Elmer).

For intra-plate variation, α1-3,4 fucosylation levels (pg/µL) were averaged for 6 replicates of human blood plasma standards (VisuCon), then standard deviation (SD) and coefficient of variation (CV) values were calculated. The averaged α1-3,4 fucosylation levels per each assay plate were used to estimate inter-plate variation, as described by SD and CV values.

RESULTS

Development of the plate assay for measurements of α1-3,4 fucosylation levels in plasma samples

The exoglycosidase plate-based assay developed in this work is based on the use of a novel α1-3,4 specific fucosidase (E1_10125), which is characterised by the ability to remove fucose residues from glycoconjugates present in blood plasma[52, 85]. Importantly, E1_10125 is capable of removing antennary fucose residues from *N*-glycans terminated with sialic acids, which is an advantage over commercially available α1-3,4 specific fucosidases. The released L-fucose residues are then subjected to an enzymatic redox reaction to produce a fluorescence signal which is directly proportional to the amount of released L-fucose residues. The released L-fucose monosaccharides are oxidised to L-fucono-1,5-lactone by a L-fucose dehydrogenase which in turn reduces NADP+ to NADPH. The NADPH is then oxidised by a diaphorase which leads to the reduction of resazurin and the formation of the fluorescent product resorufin in molar proportions stoichiometric to the released L-fucose monosaccharides (**Figure 1**).

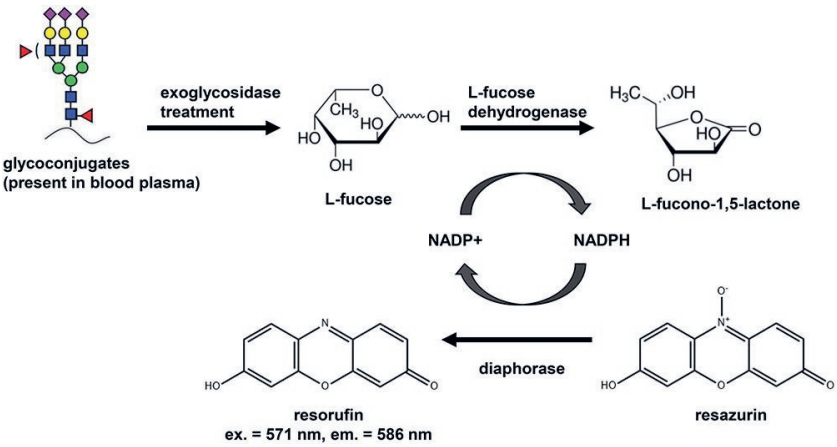


Figure 1. Schematic of the enzymatic redox reaction resulting in the formation of fluorescent product resorufin that forms the detection mechanism of the exoglycosidase plate-based assay. Resorufin measured at an excitation (ex.) wavelength of 571 nm and an emission (em.) wavelength of 586 nm. [NAPD+: nicotinamide adenine dinucleotide phosphate (oxidised); NADPH: nicotinamide adenine dinucleotide phosphate (reduced)]. This figure is available in black and white in print and in colour at *Glycobiology* online.

The fluorescence signal reaches maximum levels following 3 hours incubation and provides the best signal to background ratios. Each plasma sample was either treated with the exoglycosidase or untreated and processed in the same manner throughout the different steps of the plate-based assay. The fluorescence signal of the untreated sample was subtracted from the fluorescence signal of the exoglycosidase treated sample to exclude fluorescence background and possible interferences coming from the sample matrix. The amount of fucose in the plasma samples was then calculated based on the fucose standard curve. No fluorescence signal was generated by the E1_10125 alone, which might have influenced fucosylation level readouts and the consistency of measurements (data not shown). In this type of assay based on fluorescence measurements, readouts of low concentrations of monosaccharides might be limited by the fluorescence background signal. To be accurate, the fluorescence signal of the sample (following background subtraction) should always be greater than 0 (**Supplementary Table S1**).

Table S1. Detection of very low fucose levels in blood plasma samples might be limited by the sample background fluorescence signal. The table presents a summary of an experiment using 3 sets of plasma samples spiked with fucose standard of known concentration (in a range from 1200.00 to 0.04 pg/µL). The fucose-spiked denatured plasma samples were not exoglycosidase treated and were processed by the plate-based assay: the samples were centrifuged, the supernatants were subjected to redox reactions and the fluorescence signal measurements. Data are expressed as the mean and standard deviation (SD) of fluorescence signal together with coefficient of variation (CV) and signal to background (S/B) ratio for each concentration. S/B values below 0 (marked bold) indicate concentrations of fucose in plasma samples that were below limit of detection of the assay.

Fucose concentration in spiked plasma samples (pg/µL)	Average signal - S (U), n = 3	SD	CV	Background signal - B (U)	Average background signal – B_A (U)	S/B_A
1200.00	104152	1309	1%	34374	34833	2.99
600.00	71160	1596	2%	32480		2.04
300.00	51872	856	2%	35263		1.49
150.00	40864	1074	3%	34817		1.17
75.00	38249	1019	3%	34653		1.10
37.50	35735	1746	5%	34190		1.03
18.75	34320	1697	5%	32224		**0.99**
9.38	32343	44	0%	32193		**0.93**
4.69	34245	1360	4%	39285		**0.98**
2.34	33753	1639	5%	35227		**0.97**

1.17	35390	179	1%	37565		1.02
0.59	34515	1292	4%	35536		**0.99**
0.29	33617	1605	5%	34514		**0.97**
0.15	34145	899	3%	35346		**0.98**
0.07	33578	1613	5%	32183		**0.96**
0.04	34112	227	1%	32312		**0.98**

To determine the optimal concentration of E1_10125 during the development of the assay, 2, 4, 8 and 16 µM concentrations of the enzyme were tested on 10 µL pooled blood plasma samples. The results showed similar levels of released fucose residues irrespective of the enzyme concentration (**Supplementary Figure S1A**). Since protein concentration may vary between patient plasma samples[107], we ensured that the final concentration of E1_10125 provides maximum and consistent release of fucose regardless of protein content in plasma samples, as demonstrated on a range of pooled plasma volumes (6, 8, 10, 12 and 14 µL) (**Supplementary Figure S1B**). Based on these results, the optimal final concentration was determined to be 3 µM. Of note, in the development stage, a broader range of plasma volumes (2 – 18 µL) was tested. However, plasma volumes ≥ 16 µL led to the formation of a larger protein pellet after centrifugation, which disrupted aspiration and transfer of fucose-containing supernatants on the robotic platform, resulting in the need for manual intervention. This issue was not observed during a validation and the sample cohort analysis when using 10 µL of patients' plasma samples, which indicates that a range of 6 – 14 µL mimics the protein content variation sufficiently. Furthermore, we demonstrated that E1_10125 fucosidase at 3 µM showed high efficiency (~ 75%) in removing antennary fucose residues from intact glycoproteins present in blood plasma samples under assay conditions. This was directly compared to the efficiency of E1_10125 when applied to released *N*-glycans and showed that E1_10125 does not require released glycan substrates (**Supplementary Figure S2**).

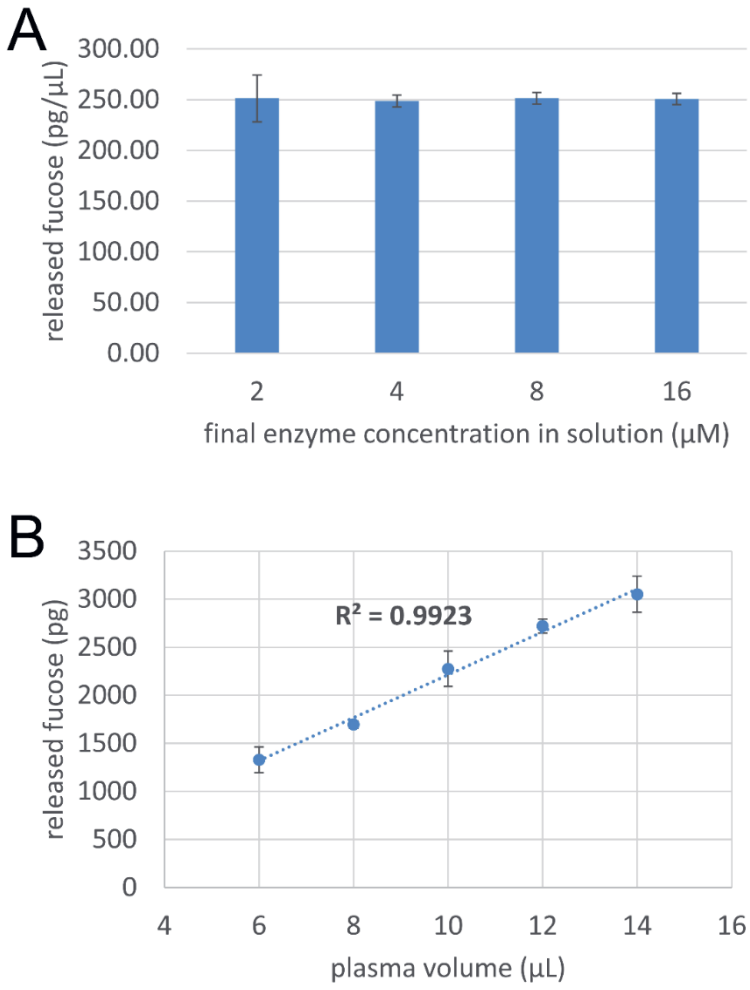
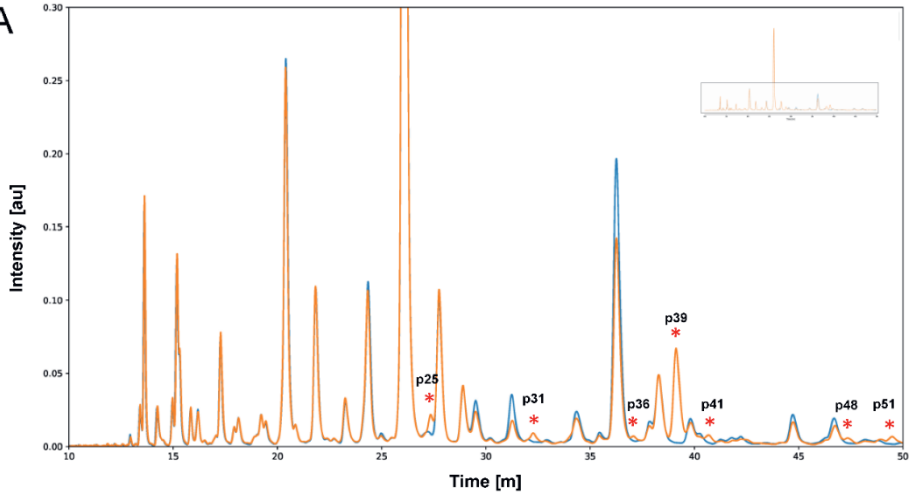
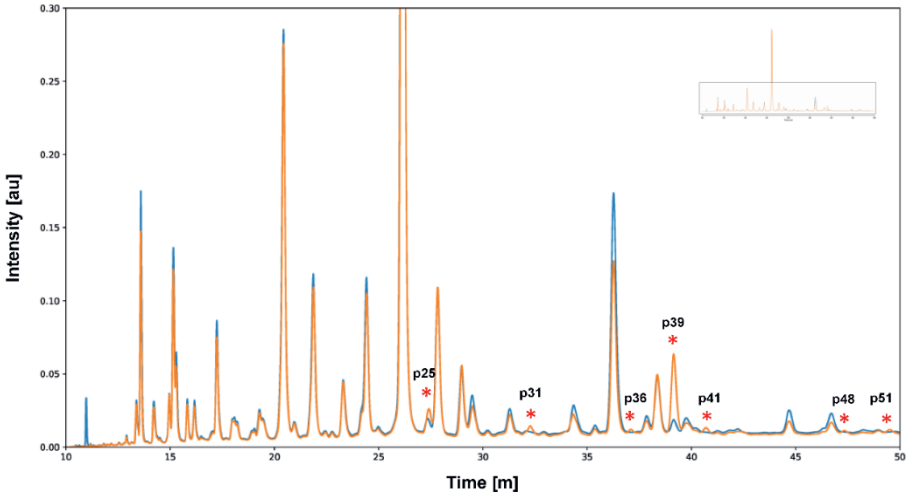


Figure S1. Performance of E1_10125 fucosidase applied at different concentrations to 10 µL plasma standard samples, n = 3 (A) and E1_10125 fucosidase applied at the final concentration of 3 µM to a range of plasma volumes, n = 3 each (B). E1_10125 provides maximum fucose release from blood plasma conjugates at the range of 2 – 16 µM final concentrations (A). 3 µM final concentration was proven to provide a good linearity (R^2 = 0.9923) for the released fucose amounts within the tested plasma volumes that mimicked the protein content variation in 10 µL of clinical blood plasma samples.

A



B



C

		p25	p31	p36	p39	p41	p48	p51	sum
Fucosidase treated plasma	set 1	0.00393	0.00128	0.00006	0.00293	0.00027	0.00042	0.00043	0.0093
	set 2	0.00409	0.00123	0.00005	0.00273	0.00029	0.00031	0.00035	0.0090
	Average	**0.00401**	**0.00125**	**0.00005**	**0.00283**	**0.00028**	**0.00036**	**0.00039**	**0.0092**
	STD	**0.00012**	**0.00004**	**0.00001**	**0.00014**	**0.00001**	**0.00008**	**0.00006**	**0.0002**
Fucosidase untreated plasma	set 1	0.00628	0.00336	0.00084	0.02336	0.00192	0.00157	0.00250	0.0398
	set 2	0.00676	0.00296	0.00075	0.01983	0.00189	0.00089	0.00134	0.0344
	Average	**0.00652**	**0.00316**	**0.00080**	**0.02160**	**0.00190**	**0.00123**	**0.00192**	**0.0371**
	STD	**0.00034**	**0.00029**	**0.00006**	**0.00249**	**0.00002**	**0.00048**	**0.00082**	**0.0038**

Figure S2. The efficiency of E1_10125 fucosidase in removing antennary fucose residues from released *N*-glycans and intact glycoproteins present in blood plasma samples. Fluorescence UHPLC chromatograms of released procainamide labeled plasma *N*-glycans derived from 8 µL pooled plasma standard (A) and plasma pellet obtained from the plate assay (B), and a table showing relative abundance of α1-3,4 fucosylated *N*-glycan peaks in full plasma *N*-glycome obtained from the plasma pellets (C). E1_10125 fucosidase digested samples (blue) were overlaid with enzymatically untreated sample chromatograms (orange). Chromatograms were normalised to the most abundant peak. α1-3,4 fucosylated glycan structure peaks are marked with asterisks (*). *N*-glycan release, exoglycosidase digestions, chromatographic separation of *N*-glycan structures and relative peak area extraction were performed as described previously[52, 84], and the content of antennary fucosylated *N*-glycans calculated for both sets of samples, each set n =2. Briefly, two sets (n = 2) of blood plasma pellets from 10 µL plasma standard (VisuCon) from a set of exoglycosidase treated and untreated samples for the plate-based assay were saved for *N*-glycan release to assess a degree of fucose release performed on denatured intact glycoproteins. Additionally, 8 µL pooled plasma standard (n = 2 sets) was used for *N*-glycan release and digestion by E1_10125 as positive control. Both sets were then compared for the degree of fucose release from intact glycoproteins vs. released *N*-glycans. The results showed that E1_10125 allows full (100%) α1-3,4 linked fucose release from released *N*-glycans whereas E1_10125 at the final concentration of 3 µM allows approximately 75% fucose release from *N*-glycans of denatured intact glycoproteins, which was calculated based on relative abundance of α1-3,4 fucosylated *N*-glycan peaks in fucosidase treated (0.0092 ± 0.0002) and untreated (0.0371 ± 0.0038) plasma samples (C).

Performance of the assay in analysing α1-3,4 fucosylation levels as a biomarker for HNF1A-MODY

Having optimised the exoglycosidase plate-based assay, 1000 blood plasma samples from the HNF1A-MODY cohort were analysed to assess α1-3,4 fucosylation levels and evaluate their diagnostic performance for the identification of cases with diabetes carrying damaging variants in the *HNF1A* gene.

Based on systematic and functional assessment of rare *HNF1A* alleles, which had been performed previously[51], the 947 participants were grouped into four *HNF1A* variant types: (likely) damaging (n = 18), (likely) benign (n = 8), variants of unknown significance (VUS, n = 5) and no *HNF1A* rare variants (n = 916). Clinical characteristics of study participants are summarised in **Table 1**.

Table 1. Clinical characteristics of study participants

	(Likely) damaging allele	(Likely) benign n = 8 cases	Variants of unknown significance (VUS)	No rare *HNF1A* allele variant
Sex, male (n)	5	4	2	529
Age at recruitment (years)	40 ± 17	48 ± 11	54 ± 14	47 ± 11
Age at diagnosis (years)	26 ± 9	36 ± 6	37 ± 10	35 ± 7
Diabetes duration (years)	13 ± 12	9 ± 7	17 ± 11	12 ± 10
BMI (kg/m^2)	26 ± 5	35 ± 5	28 ± 6	31 ± 7
hsCRP (mg/L)	0.8 ± 1.3	5.2 ± 5.7	4.0 ± 6.3	6.2 ± 29.7
HbA1c (%)	7.7 ± 1.7	8.8 ± 2.1	6.7 ± 1.3	7.9 ± 2.7
C-peptide (nmol/L)	0.42 ± 0.21	0.81 ± 0.62	0.60 ± 0.57	5.11 ± 71.47
Total cholesterol (mmol/L)	4.75 ± 1.09	5.07 ± 0.88	4.74 ± 1.21	4.69 ± 1.23
HDL (mmol/L)	1.4 ± 0.3	1.1 ± 0.2	1.3 ± 0.4	1.3 ± 3.4
Triglycerides (mmol/L)	1.2 ± 0.5	2.0 ± 1.0	1.1 ± 0.3	2.0 ± 1.7

Mean values are presented for each group (± standard deviation)

From the 1000 samples tested, 947 showed a background to fluorescence signal > 0 and blood plasma α1-3,4 fucosylation levels from these samples were subsequently subjected to statistical analyses.

Box plot analysis was used to evaluate differences in α1-3,4 fucosylation levels between the four *HNF1A* variant groups. The analysis showed significant differences ($p \leq 0.05$) in α1-3,4 fucosylation levels between groups of (likely) damaging vs. (likely) benign and (likely) damaging vs. cases without rare *HNF1A* variants (**Figure 2**). No significant difference in α1-3,4 fucosylation levels was observed between (likely) benign and a group of cases without rare *HNF1A* variant groups. Data outliers were identified within each group, 1 case within (likely) damaging group and 24 cases within cases without *HNF1A* mutation and removed prior to the further analyses[89]. The average fucose content in blood plasma for each defined group was as follows: 79.26 ± 35.63 pg/µL for (likely) damaging, 241.01 ± 125.41 pg/µL for (likely) benign, 185.71 ± 111.15 pg/µL for VUS and 193.60 ± 99.58 pg/µL for cases without rare *HNF1A* variants.

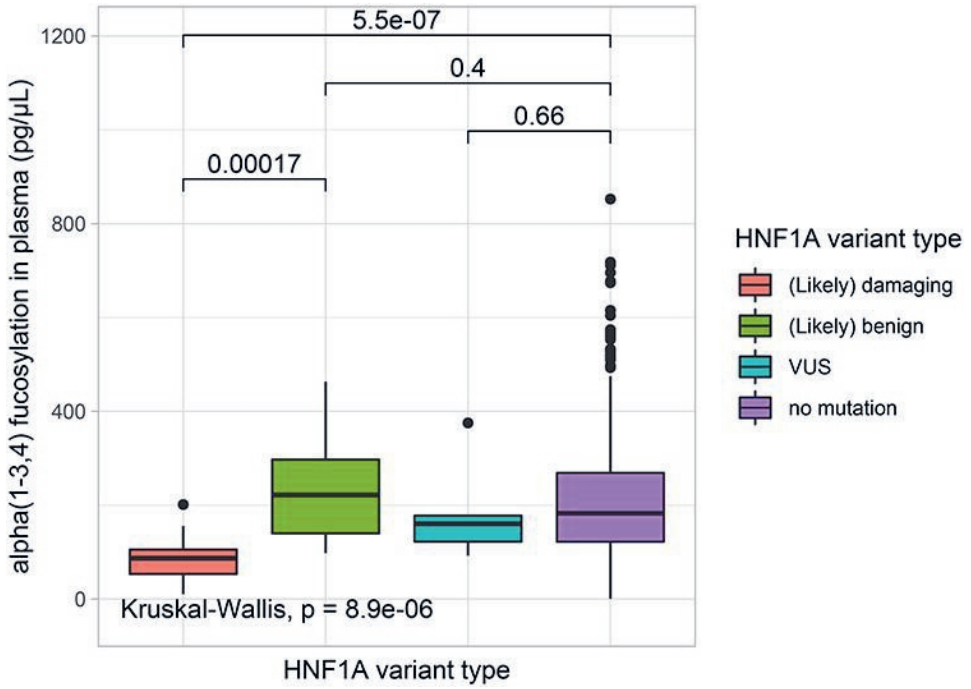


Figure 2. Box plots presenting differences in α1-3,4 fucosylation levels in blood plasma (pg/µL) measured using the exoglycosidase plate-based assay for groups of patients with different *HNF1A* variant groups. Each box represents 25th–75th percentile, the median is marked by a vertical line, whiskers indicate values that are within 1.5 × IQR of the hinge. Outliers are displayed as black filled circles. The lines and numbers above the box plots indicate the *P*-value when comparing two categories using The Wilcoxon–Mann–Whitney test. The analysis with $P \leq 0.05$ is considered statistically significant.

Next, receiver operating characteristics (ROC) curves were used to estimate the differentiating power of α1-3,4 fucosylation levels to identify patients with HNF1A-MODY carrying damaging variants in the *HNF1A* gene. The ROC analysis showed that α1-3,4 fucosylation levels provide good differentiating power between a group of (likely) damaging cases and the group of individuals without rare *HNF1A* variant , as determined by the measurement of the area under the curve (AUC) that was found to be 0.87 with 94% sensitivity, 71 % specificity at the cut-off of 129.43 pg/µL (**Figure 3A**). Additionally, the diagnostic performance of α1-3,4 fucosylation levels was tested for (likely) damaging vs. (likely) benign variant groups giving an AUC of 0.95 with 94% sensitivity, 75% specificity at the cutoff of 137.59 pg/µL (**Figure 3B**). By applying α1-3,4

fucosylation levels as a biomarker for HNF1A-MODY, two cases from the VUS group would have been classified as carrying damaging mutations.

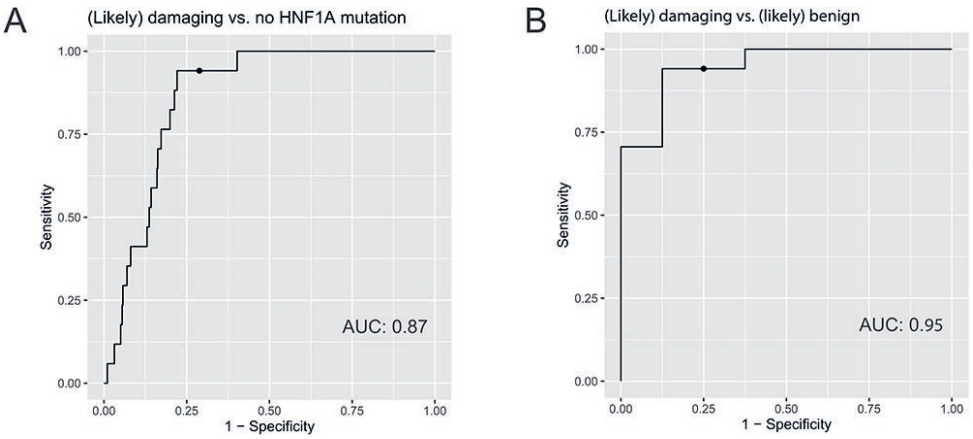


Figure 3. ROC curves illustrating the performance of α1-3,4 fucosylation levels in differentiating cases with (likely) damaging vs. no *HNF1A* variants (A) and (likely) damaging vs. (likely) benign variants in the *HNF1A* gene (B). The AUC values are displayed for each ROC curve. The optimal cut-off points are displayed as a dot on the precision recall curve for each ROC curve.

Intra- and inter-assay variations for measurements of α1-3,4 fucosylation levels in blood plasma samples using the exoglycosidase plate-based assay

In order to evaluate the effectiveness and robustness of the exoglycosidase plate-based assay, 72 plasma standard samples, were distributed over 12 assay plates measured on different days. The results showed good repeatability and precision for measurements of α1-3,4 fucosylation levels in blood plasma samples with CVs of 9% for average intra-plate variation and 10% for inter-plate variation (**Table 2**).

Associations between α1-3,4 fucosylation levels, age, sex and clinical markers

We investigated associations between α1-3,4 fucosylation levels and either sex or age within individuals with diabetes without rare variants in the *HNF1A* gene. A significant association between α1-3,4 fucosylation levels and sex (regression coefficient $\beta = 0.25$, $\rho = 0.0005$) but not age was observed. Average α1-3,4 fucosylation levels were found to be 180.9 ± 97.8 pg/µL in females (n = 362) and 204.2 ± 100.2 pg/µL in males (n = 510).

Table 2. Intra- and inter-assay variations for measurements of α1-3,4 fucosylation levels in blood plasma samples by the exoglycosidase plate-based assay. Intra- and inter-assay variations are described by coefficient of variation (CV) values that were calculated based on the amount of released fucose for human blood plasma standards (n = 6) distributed over 12 assay plates.

Intra-plate variation			
Plate	Average fucose level in plasma (pg/µL), n = 6	SD	CV
1	235.65	13.11	6%
2	213.89	55.97	26%
3	224.09	30.18	13%
4	200.53	10.62	5%
5	250.55	6.62	3%
6	219.50	17.17	8%
7	242.71	10.95	5%
8	213.80	27.16	13%
9	256.58	7.85	3%
10	230.28	8.42	4%
11	175.92	19.79	11%
12	220.35	22.09	10%
Average intra-plate			9%
Inter-plate variation	223.65	22.14	10%

Previously, we reported significant correlations between CRP levels and antennary fucosylation levels, which were measured using antennary fucosylated *N*-glycan traits by an LC-based method[52]. Here, we tested α1-3,4 fucosylation levels against a panel of inflammatory and metabolic markers (**Table 1**). We found a weak correlation between α1-3,4 fucosylation and CRP levels (correlation coefficient $r = 0.29$, $p < 2.2 \times 10^{-16}$) within the unselected sample cohort of young-adult onset diabetes (**Supplementary Figure S3A**). Another weak correlation was found for α1-3,4 fucosylation levels and body mass index (BMI) with a correlation coefficient r = 0.17, p < 1.2 x 10-6 (**Supplementary Figure S3B**) in the same cohort.

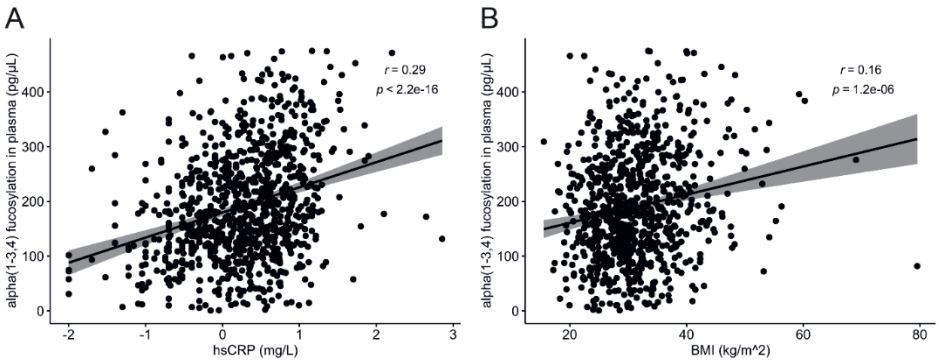


Figure S3. Correlations between CRP and α1-3,4 fucosylation levels (A), and BMI and α1-3,4 fucosylation levels (B) observed within the sample cohort. hsCRP concentrations were logarithmically transformed.

We next compared the fucosylation indexes previously measured using an LC-based method[52], which reflect levels of antennary fucosylation in the full plasma *N*-glycome, with the α1-3,4 fucosylation levels in blood plasma samples measured by the exoglycosidase plate-based assay. The analysis was carried out for 31 cases that had been diagnosed with mutations in the *HNF1A* gene, which included groups of (likely) damaging, (likely) benign variants and VUS. The Spearman's correlation analysis showed a strong correlation between the fucosylation indexes and α1-3,4 fucosylation levels in blood plasma samples with a correlation coefficient r = 0.85 (**Figure 4A**). When considering all previously analysed 320 diabetes cases with and without mutations in the *HNF1A* gene, a significant but weaker correlation with the correlation coefficient r = 0.68 was obtained (**Figure 4B**)

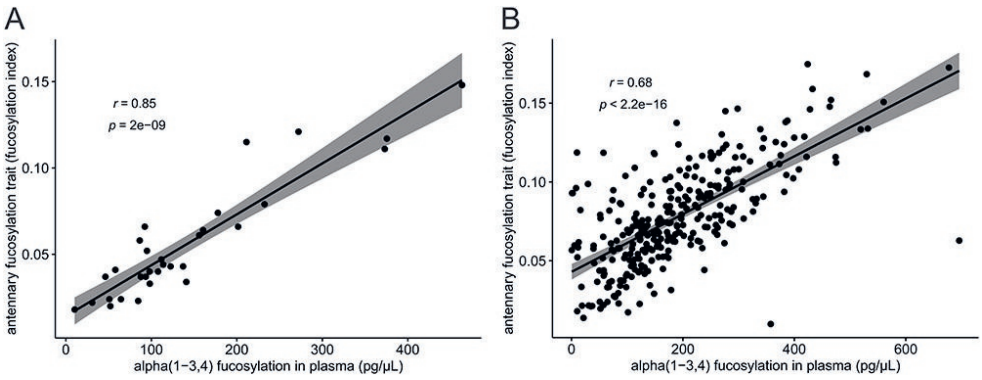


Figure 4. Correlations between *N*-glycan antennary fucosylation levels measured as indexes by LC method in the previous study (Demus et al. 2021[52]) and α1-3,4 fucosylation levels measured by the exoglycosidase plate-based assay for 31 HNF1A-MODY positive cases (A), and 320 diabetes cases including 31 HNF1A-MODY positive and 289 negative cases (B).

DISCUSSION

Diagnosis of HNF1A-MODY is challenging as current clinical criteria to guide genetic screening for MODY have been proven insensitive[108] and increasingly non-specific[109]. Overlapping of clinical features of MODY with other types of diabetes, poor access and the cost of genetic testing often leads to misdiagnosis of MODY patients and incorrect disease management[109]. Since the GWAS study from 2010 revealed that damaging variants in the *HNF1A* gene are associated with decreased levels of antennary fucosylated *N*-glycans[82], this potential glycan biomarker for HNF1A-MODY has been studied using a combination of fluorescently labelled glycans and LC separation techniques. Its diagnostic accuracy has been tested on large cohorts of cases with diabetes carrying different variants in the *HNF1A* gene. Individuals with damaging alleles in the *HNF1A* gene are known to have malfunctioning molecular mechanisms in which HNF1α factor is involved. The HNF1α is a master regulator of several genes, among others, genes controlling β-cell function and growth[74] and genes encoding fucosyltrasferases[82]. Therefore, individuals that carry damaging variants in the *HNF1A* gene represent decreased levels of α1-3,4 fucosylation. The excellent differentiating performance of blood plasma antennary fucosylated *N*-glycans as a biomarkers for HNF1A-MODY with AUCs of over 0.90 has been reported[51-53]. By applying those glycan biomarkers, it was possible to identify rare cases with damaging variants in the *HNF1A* gene. However, this analytical method has not been translated into a widely available diagnostic tool.

As a response to the current demand for a simple and effective diagnostic tool for the detection of this well-studied *N*-glycan biomarker for HNF1A-MODY, we have developed the exoglycosidase plate-based assay with fluorescence output to measure α1-3,4 fucosylation levels in blood plasma samples. This assay employs a novel α1-3,4 specific fucosidase (E_10125), which is capable of releasing fucose residues from glycoconjugates present in blood plasma samples[85]. This fucosidase works efficiently on *N*-glycan structures terminated with sialic acids and shows high efficiency (~ 75%) in removing *N*-glycan antennary fucose residues from intact glycoproteins present in

plasma samples. This advantage has been exploited in this study. The remaining portion (~ 25%) of undigested fucose from *N*-glycans, when applying E1_10125 fucosidase to denatured plasma samples, might be due to the inaccessibility of fucose residue substrates to the enzyme in this case where the glycans remain attached to the proteins. Nevertheless, the plate-based assay targets α1-3,4-linked fucose from all glycoconjugates (*N*- and *O*-glycosylated proteins and glycolipids) present in plasma samples and, as demonstrated by the consistency and excellent linearity of released fucose levels, the undigested portion of *N*-glycan fucose residues does not interfere with the results.

The validity of the assay was tested on 1000 blood plasma samples from a diabetes cohort, which included diagnosed cases with and without rare variants in the *HNF1A* gene. The ROC analysis showed that α1-3,4 fucosylation levels assessed from blood plasma samples by the exoglycosidase plate-based assay provide very good discriminatory power to identify cases with damaging variants in the *HNF1A* gene, as tested against a group of cases with benign (AUC = 0.95) and no rare variants (AUC = 0.87) in the *HNF1A* gene. The results for cases with damaging vs. no rare variants in *HNF1A* are similar to the AUCs of 0.90 for two single glycans structures reported using a LC-based method from the same sample cohort[51]. We previously determined an antennary fucosylation trait, which reflects overall changes in antennary fucosylation in blood plasma *N*-glycome[52]. Here, *N*-glycan antennary fucosylation levels measured as indexes by applying this derived antennary fucosylation trait were correlated with α1-3,4 fucosylation levels of blood plasma glycoconjugates measured by the exoglycosidase plate-based assay for 31 HNF1A-MODY cases and all 320 cases with young-adult onset non-autoimmune diabetes, separately. The respective Spearman's correlation coefficients of 0.85 and 0.68 obtained in the correlation analysis are considered relatively low and might suggest varying levels of α1-3,4-linked fucose residues released from *O*-glycans and glycolipids present in blood plasma. Higher values for the correlation coefficient would have been expected in the case of proportionate amounts of α1-3,4-linked fucose derived from *N*-glycans and other

glycoconjugtates in plasma samples. Nevertheless, the significance of the correlation analysis together with results obtained for the intra- and inter-plate variations from the current cohort study demonstrate the excellent repeatability of measurements and the validity of the exoglycosidase plate-based assay.

The performance of a biomarker based on α1-3,4 fucosylation might be influenced by inflammatory events that affect glycosylation profiles[52, 98]. A large proportion (~ 50%) of *N*-glycans containing α1-3,4 fucosylation derives from acute phase proteins expressed during inflammation[94] and, here, we have confirmed that α1-3,4 fucosylation levels weakly correlate with CRP levels (correlation coefficient r = 0.29) that are indicative of inflammation as well as a non-alcoholic fatty liver disease related to participants BMI[99]. For the later, overexpression of fucosyltransferase FUT6 has been reported, which might lead to altered expression of other fucosyltrasferases due to their interconnection with the HNF1α transcription factor and the availability of guanosine diphosphate-fucose (GDP-fucose) donor substrate[82, 110]. Furthermore, the weak correlation between α1-3,4 fucosylation levels and BMI (correlation coefficient r = 0.17), which has also been observed within the current sample cohort, might be explained by the obesity-linked pro-inflammatory state, and secretion of inflammatory mediators as well as CRP[111]. In addition, we found that the α1-3,4 fucosylation levels were associated with sex, but not age, with higher α1-3,4 fucosylation levels in male individuals, consistent with previous findings reporting that fucosylation is gender dependent[112]. Together this data indicates that different cut-off values for α1-3,4 fucosylation levels will need to be applied to men and women in a real diagnostic setting.

CRP, which is already widely used in clinical testing, has been previously evaluated as a biomarker for HNF1A-MODY and provided a good diagnostic performance with an AUC of 0.83 (88% sensitivity and 69% specificity at the cutoff 0.81 mg/L), although not as performant as antennary fucosylated *N*-glycan and α1-3,4 fucosylation biomarkers[51]. In addition, the application of CRP as a specific biomarker for HNF1A-MODY is limited

as CRP levels rise rapidly during inflammation[113] which may confound the results. The influence of inflammation on the performance of the glycan biomarker for HNF1A-MODY is possible yet still less significant than in the case of CRP. Altogether, the good performance of α1-3,4 fucosylation as a biomarker for HNF1A-MODY implies that employing easily accessible technology, such as the described exoglycosidase plate-based assay, could serve as a screening step to identify high risk cases and select individuals for the *HNF1A* sequencing which is the current diagnostic gold standard test. This is likely to be cost-saving by narrowing the number of cases requiring expensive genetic testing.

The sequences of *FUT3*, *FUT5*, and *FUT6* are highly polymorphic and the presence of single-nucleotide polymorphisms (SNPs) in genes encoding fucosyltransferases might affect α1-3,4 fucosylation levels measured by the plate-based assay[114]. Genetic variations can lead to inactivation of fucosyltrasferases, for example, inactive FUT3 in individuals with Lewis negative blood type[115]. Information on the presence of SNPs in genes encoding fucosyltransferases was not available in the clinical data of the sample cohort used in this study. Therefore, we are not able to estimate the significance of occurrence of these SNPs on α1-3,4 fucosylation levels measured by the plate-based assay. Exclusion of individuals with loss of function SNPs in fucosyltransferase genes might lead to further improvement of the AUC and different cut-off values for α1-3,4 fucosylation biomarker. We consider this aspect as a limitation of the current study with a potential for future research and investigation.

There are several advantages to the exoglycosidase plate-based assay over other methods used for glycosylation analysis. First, the exoglycosidase plate-based assay allows timely and cost-effective high-throughput screening of large numbers of plasma samples. Currently, the LC-based approach requires extensive sample preparations and a long chromatographic separation of fluorescently-tagged glycans, often exceeding 1 hour per sample. In comparison, our exoglycosidase plate-based assay provides semi-automated sample preparation using the robotic platform followed by readouts of

fluorescence signals that allow the assessment of absolute α1-3,4 fucosylation levels from 96 plasma samples within 24 hours. Furthermore, the assay uses a simple technology based on fluorescence readouts and generates results that are easy to process and interpret, with no need for high-end instrumentation or expertise. This type of exoglycosidase assay has a potential to become an alternative to immuno-/lectin-based biochemical assays, the use of which might be limited by the availability and binding affinity of the detection antibodies and lectins[116]. It is expected that advances in glycosidase discovery will enhance a broader application of the exoglycosidase plate-based assay and enable the measurement of absolute levels of other monosaccharides in various types of samples, including released glycans, intact proteins, biopharmaceuticals and clinical samples[85]. In conclusion, the exoglycosidase plate-based assay developed in this work enables robust and high-throughput screening of absolute α1-3,4 fucosylation levels in blood plasma samples. The cohort study confirmed the excellent performance of α1-3,4 fucosylation levels as a clinical biomarker for HNF1A-MODY, which allows identification and classification of cases with diabetes carrying damaging variants in the *HNF1A* gene. The results of this work should facilitate the translation of this glycan biomarker into clinical practice and the development of a clinically relevant, widely available diagnostic test.

Acknowledgments

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 722095. The authors gratefully acknowledge the support of the Biotechnology and Biological Sciences Research Council (BBSRC) with contribution from the Innovate UK Biocatalyst grant Glycoenzymes for Bioindustries (BB/M029042/) and the BBSRC Institute Strategic Programme Grant Gut Microbes and Health BB/R012490/1 and its constituent project BBS/E/F/000PR10356 (Theme 3, Modulation of the gut microbes to promote health throughout life). The research was supported by the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre (BRC). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. AJ was a Diabetes UK George Alberti fellow during the research.

Authors would like to thank Archana Mili Shubhakar (Ludger) for the technical expertise in the assay automation. Matthew Doherty (Ludger) for help in creating figures. Osmond Rebello (former Ludger) for working simultaneously on the development of glycosidase plated-based assays.

The CroDiab study was approved by Ethics Committee of Clinical Hospital Merkur and University of Zagreb, Faculty of Pharmacy and Biochemistry, Croatia. The study was conducted in accordance with the Declaration of Helsinki. The YDX study was approved by the Oxfordshire Local Research Ethics Committee, and all subjects gave written informed consent.

CHAPTER 4

4. Large-scale analysis of apolipoprotein CIII glycosylation by ultrahigh resolution mass spectrometry

Daniel Demus[1,2], Annemieke Naber[3], Viktoria Dotz[1†], Bas C. Jansen[1,2], Marco R. Bladergroen[1], Jan Nouta[1], Eric J. G. Sijbrands[3], Mandy van Hoek[3], Simone Nicolardi[1], Manfred Wuhrer[1]

[1] Leiden University Medical Center, Center for Proteomics and Metabolomics, Leiden, Netherlands

[2] Ludger Ltd., Culham Science Centre, Abingdon, Oxfordshire, United Kingdom

[3] Department of Internal Medicine, Erasmus University Medical Center, Rotterdam, Netherlands

Reprinted and adapted: Demus D, Naber A, Dotz V, Jansen BC, Bladergroen MR, Nouta J, Sijbrands EJG, Van Hoek M, Nicolardi S, Wuhrer M. Large-Scale Analysis of Apolipoprotein CIII Glycosylation by Ultrahigh Resolution Mass Spectrometry. Front Chem. 2021 May 7;9:678883. doi: 10.3389/fchem.2021.678883

ABSTRACT

Apolipoprotein-CIII (apo-CIII) is a glycoprotein involved in lipid metabolism and its levels are associated with cardiovascular disease risk. Apo-CIII sialylation is associated with improved plasma triglyceride levels and its glycosylation may have an effect on the clearance of triglyceride-rich lipoproteins by directing these particles to different metabolic pathways. Large-scale sample cohort studies are required to fully elucidate the role of apo-CIII glycosylation in lipid metabolism and associated cardiovascular disease. In this study, we revisited a high-throughput workflow for the analysis of intact apo-CIII by ultrahigh-resolution MALDI FT-ICR MS. The workflow includes a chemical oxidation step to reduce methionine oxidation heterogeneity and spectrum complexity. Sinapinic acid matrix was used to minimize the loss of sialic acids upon MALDI. MassyTools software was used to standardize and automate MS data processing and quality control. This method was applied on 771 plasma samples from individuals without diabetes allowing for an evaluation of the expression levels of apo-CIII glycoforms against a panel of lipid biomarkers demonstrating the validity of the method. Our study supports the hypothesis that triglyceride clearance may be regulated, or at least strongly influenced by apo-CIII sialylation. Interestingly, the association of apo-CIII glycoforms with triglyceride levels was found to be largely independent of body mass index. Due to its precision and throughput, the new workflow will allow studying the role of apo-CIII in the regulation of lipid metabolism in various disease settings.

INTRODUCTION

Lipid metabolism is regulated by complex biological mechanisms in which apolipoproteins – proteins embedded in lipoprotein particles – modulate the transport and availability of blood lipids[117]. Apolipoprotein-CIII (apo-CIII) is a 79 amino acid glycoprotein present on the surface of triglyceride-rich lipoproteins and is an inhibitor of lipoprotein lipase (LPL), an enzyme that hydrolyses triglycerides into fatty acids[118-120]. Apo-CIII has been associated with increased monocyte adhesion to the endothelium[121] and enhanced binding of apoB-containing lipoproteins to vascular proteoglycans[122]. High apo-CIII levels are associated with hypertriglyceridemia[123-125] and increased cardiovascular disease risk in the general population[123, 126, 127] and diabetes mellitus[128, 129]. Recently, the clinical interest for this protein has increased due to the promising results obtained from antisense oligonucleotide-based therapies for the reduction of apo-CIII and triglyceride levels[125, 130-132].

Apo-CIII exists in four major proteoforms: one non-glycosylated form (apo-$CIII_{0a}$) and three *O*-glycosylated variants with a core 1 (T-antigen) glycan structure, which is either non-sialylated (apo-$CIII_{0c}$), monosialylated (apo-$CIII_1$) or disialylated (apo-$CIII_2$)[133, 134]. Low-abundance fucosylated, non-sialylated apo-CIII forms have also been described[35]. It has been shown that not only the levels of apo-CIII but also the specific glycoforms and their relative expression control triglyceride metabolism[57, 58]. For example, an inverse association between apo-$CIII_2$/apo-$CIII_1$ ratio and triglyceride levels has been confirmed by two independent studies[57, 59]. It has also been shown that sialylation modulates the apo-CIII affinity for hepatic receptors that clear lipoprotein particles[59] and that different proteoforms of apo-CIII may affect the inhibition of LPL[135] and the interaction of LDL with the vascular wall[136]. Since the association of different apo-CIII proteoforms with specific cardiometabolic endpoints has not been fully elucidated, further research in large sample cohorts is warranted.

We have developed a high-throughput method based on magnetic-bead extraction and matrix-assisted laser desorption/ionization (MALDI) and ultrahigh-resolution Fourier

transform ion cyclotron resonance (FT-ICR) mass spectrometry (MS) for the analysis of serum apo-CIII proteoforms[35, 137]. Apo-CIII contains methionine residues, which can be (partially) oxidised during biological processes *in vivo*[138], sample processing and freeze-thaw cycles[139]. The presence of different oxidoforms increases mass spectra complexity, which complicates MS data processing and affects the repeatability of measurements. Although the analyte oxidation may not pose a serious challenge in MALDI MS analysis of single samples, it can seriously impact the precision and accuracy of quantitative measurements in large sample cohorts.

In the current study, we have applied a modified workflow employing a previously established MALDI FT-ICR MS method preceded by a chemical oxidation step for complete oxidation of apo-CIII methionine residues. This results in highly reproducible high-throughput measurements for relative quantification of apo-CIII proteoforms in a large number of plasma samples varying in protein oxidation levels. Furthermore, we have adopted sinapinic acid (SPA) as a MALDI matrix to minimize the loss of sialic acid induced by MALDI. The high-throughput quantitation software, MassyTools[140], was here further developed to facilitate semi-automated MS data processing for intact proteins. The validity of the new workflow was tested on a clinical cohort comprised of 771 plasma samples, which allowed the evaluation of the relationship between apo-CIII glycoforms and metabolic biomarkers, such as BMI, cholesterol, and triglyceride levels.

MATERIALS AND METHODS

Clinical samples

Blood plasma samples from a group of individuals without diabetes of the DiaGene Study were used. The DiaGene Study is a case-control study comprising 1886 type-2 diabetes patients and 854 controls without diabetes, from the areas of Eindhoven and Veldhoven, in the Netherlands. The study is described in detail elsewhere[141]. For the current study, after quality control, apo-CIII glycosylation data were available for 771 samples, in 746 whereof, data on clinical characteristics were available. All participants gave their written informed consent. This study was approved by the Medical Ethics Committees of the Erasmus University Medical Center, Catharina Hospital and Maxima Medical Center.

Clinical information and blood samples were obtained at baseline, as described previously[141]. Triglycerides and cholesterol concentrations were measured using standard clinical chemistry essays and reported by the collecting clinic. Non-high-density lipoprotein (non-HDL)-cholesterol was calculated by subtracting the high-density lipoprotein (HDL)-cholesterol from the total cholesterol, body mass index (BMI) was calculated by dividing the body mass (in kg) by the square of the body length (in m). Triglyceride concentrations were logarithmically transformed before linear regression analysis, because of non-normal distribution.

Chemicals

Magnetic beads (Dynabeads RPC-18) were purchased from Invitrogen Dynal AS, Oslo, Norway. VisuCon-F plasma standard from Affinity Biologicals, Ancaster, Canada. Hydrogen peroxide 30%, ethanol and acetone were purchased from Merck, Darmstadt, Germany. Acetonitrile (ACN) was from Biosolve Chimie SARL, France. Trifluoroacetic acid (TFA) was purchased from Thermo Fisher Scientific, Tewksbury, MA. Sinapinic acid (SPA) and α-Cyano-4-hydroxycinnamic acid (HCCA) from Sigma-Aldrich. Ultrapure milliQ water (18 MΩ·cm at 25°C) was used throughout.

High-throughput RP-C18 solid-phase extraction (SPE) of plasma proteins

Plasma standards (VisuCon-F) were randomised over cohort sample plates. 10 µL of human blood plasma was transferred from the cohort sample plates into 96-well skirted PCR plates (4ti-0960/C, 4titude, Dorking, UK). 15 µL of an oxidizing solution (12% H_2O_2/0.5% TFA in water) was added to each sample. The plate was sealed with a pierce foil seal (4ti-0521, 4titude Ltd, Wotton, Surrey, UK) and incubated for 1 hour at 37°C. Subsequently, the plate was cooled at 4°C for 30 min and centrifuged briefly at 800 × *g*. The pierce foil was removed and the plate was transferred onto a liquid handling robot (Hamilton, Bonaduz, Switzerland) where solid-phase extraction (SPE) was carried out as follows: the RP-C18 beads were activated by three washes using acetonitrile (ACN) and trifluoroacetic acid (TFA) solution in water (first wash using 50% ACN/0.1% TFA followed by two washes with 0.1% TFA). Next, plasma samples were transferred to the activated beads and incubated for 10 min at room temperature. The incubation was followed by three washes: one wash using 15% ACN and two washes with 0.1% TFA. Proteins were eluted by adding 15 µL of 50% ACN/0.1% TFA in water and incubating for 5 min at room temperature. For MALDI spotting, 2 µL of sample eluates were mixed with either 16 µL of sinapinic acid solution (1.3 g/L in 2:1 v/v ethanol/acetone) or 15 µL alpha-cyano-4-hydroxycinnamic acid solution (1.4 g/L in 2:1 v/v ethanol/acetone). 1.5 µL of each sample mix was spotted in duplicate onto a MALDI AnchorChip target plate (800 µm anchor diameter; Bruker Daltonics, Bremen, Germany) and allowed to air-dry before MALDI MS analysis.

MALDI FT-ICR mass spectrometry and MS data analysis

All MALDI MS experiments were performed on a 15 T solariX XR FT-ICR mass spectrometer (Bruker Daltonics) equipped with a Smartbeam II™ laser system (355 nm wavelength) and a ParaCell detector. All spectra were acquired in the *m/z*-range 3495-30000, from the average of ten scans of 200 laser shots (at 500 Hz) each using 524288 data points. The analyser parameters were set as previously reported[142]. Briefly, measurements were performed with high trapping potentials (up to 8.5 V) and high ParaCell DC biases (up to 8.8 V) and with a Sweep excitation power of 57% for 13.5 µs.

A laser power of 20% and "medium" laser focus was used for MS measurements using HCCA, while a laser power of 30% and "ultra-large" focus was used with SPA.

MS data was converted as *.xy files using DataAnalysis (Bruker, ver. 5.0 SR1) before subsequent processing by MassyTools[140]. Originally developed for automated processing of MALDI MS data from released *N*-glycans, MassyTools was here adjusted to large molecules. Specifically, the modification entailed that the full isotopic envelope is calculated for each molecule, from which the largest contributing isotopes are selected for quantitation until a user-specified threshold (v1.0.2-alpha b180626a). The processing parameters listed below are explained in detail in work by Jansen et al.[140]. All spectra were internally calibrated using the most abundant isotopic peak of oxidoforms of the four major apo-CIII proteoforms (apo-CIII0a+2Ox, apo-CIII0c+2Ox, apo-CIII1+2Ox and apo-CIII2+2Ox; m/z values of 8797.228, 9162.361, 9453.456 and 9744.552, respectively). Data curation was performed in a semi-automated manner using Microsoft Excel and a custom-made script in RStudio (version 1.1.463). Spectra were excluded from further analysis if either calibration by MassyTools failed or if their "Fraction of Analyte Area - Background Area above S/N cut-off" parameter value was below 3*IQR of the mean value). Excluded spectra were spot-checked by visual inspection to confirm their low quality, especially in case of spectra which were close to the cutoff. Out of each sample spot duplicate, the spectrum with the higher absolute area sum of the four proteoforms was kept for further processing. All analytes of interest had to pass set cut-off values for minimum two quality parameters: signal-to-noise ratio (S/N) of the most-abundant isotopic peak, mass accuracy (MME) and isotopic pattern quality (IPQ). The cut-off values were as follows: S/N ≥ 9, MME ≤ 10 and IPQ ≤ 0.2. Spectra, in which an analyte met two out of three quality parameters, were subjected to additional manual inspection for the presence of interfering species. Peak intensities were measured as peak areas of the most abundant isotopic peak. MassyTools extraction parameters were chosen experimentally for optimal data extraction and were as follows: peak extraction width = 0.3, minimum fraction of total isotopic distribution used for extraction = 0.5, background detection window = 20.

Intra- and inter-plate repeatability were calculated based on 136 plasma standards (VisuCon) distributed over 31 sample plates. For intra-plate repeatability, the mean values of relative peak intensities of apo-CIII0a+2Ox, apo-CIII0c+2Ox, apo-CIII1+2Ox and apo-CIII2+2Ox, standard deviations (SD) and coefficients of variations (CV) were calculated for the standards per sample plate. To assess average intra-plate repeatability, the per-plate CVs were averaged. The per-plate means were averaged, SDs and CVs were calculated over all 31 plates to estimate the inter-plate repeatability. Statistical analyses were performed in SPSS version 25 and data visualization was performed in RStudio (version 1.1.463) and Microsoft Excel. To compare apo-CIII glycosylation in males and females, an independent-samples T-test was performed. The distribution of clinical variables was considered normal when Skewness and Kurtosis were within the range of -1 to +1. The association of apo-CIII glycoforms with age, BMI, HDL-cholesterol, non-HDL-cholesterol, low-density lipoprotein (LDL)-cholesterol, triglycerides and total cholesterol was analysed using univariate linear regression models, where apo-CIII glycoforms were entered as dependent variables. Triglycerides were logarithmically transformed prior to analysis, because of a non-normal distribution, to derive p-values and R-square. To investigate the influence of BMI on the association of apo-CIII glycosylation with triglyceride levels, a multiple linear regression model was used which included BMI and triglycerides as independent variables. $P < 0.05$ was considered statistically significant.

RESULTS AND DISCUSSION

The controlled oxidation of methionine residues reduces the complexity of mass spectra

A common event observed in proteomics studies is the oxidation of methionine residues due to biological and pathological processes occurring *in vivo*[138], sample storage and processing[139]. These reactions are so common that, in bottom-up studies, methionine oxidation is often included in database searches as a variable or even fixed modification. However, in general, peptides generated by enzymatic digestions (e.g. using trypsin) are small and often do not contain methionine residues and although a (partial) oxidation of methionine residues increases the number of peptides in a digest, these unwanted reactions do not significantly affect the analysis[143-145]. In a clinical setting, the use of fresh samples may be an ideal approach, especially for diagnostic purposes based on profiling of intact protein. Whereas in cohort studies, the collection of large numbers of clinical samples, their storage, transfer between institutions and multiple use may lead to oxidation processes that affect the analysis of intact proteins by increasing the heterogeneity of proteoforms detected in a spectrum. The higher complexity increases the chance of overlapping signals and reduces the sensitivity of measurements due to the spreading of the signal over a higher number of species. This results in MS spectra characterised by the presence of interfering species and very low abundant analyte peaks, which do not meet acceptable spectral quality criteria for consideration in the statistically significant quantitative analysis.

Apo-CIII contains two methionine residues which can be oxidised to form methionine sulfoxide (MetO) and methionine sulfone (MetO2) although this latter form requires harsher oxidizing conditions[146, 147]. Previously, MALDI-TOF MS methods have been used in analyses of apo-CIII proteoforms[148]. In such low-resolution methods, apo-CIII oxidoforms cannot be resolved, however, their presence results in the broadening and distortion of apo-CIII proteoforms' signals, which can eventually overlap or interfere with signals of other proteins affecting their quantification. The chance of signal interference increases when SPE is used for the enrichment of apo-CIII as it leads to the

co-enrichment of other small plasma proteins. The application of more specific enrichment methods, such as immunocapture, may help to reduce signals interfering with the various apo-CIII oxidoforms, but was not implemented in the present study for simplicity reasons. Of note, apo-CIII proteoforms have been analysed by methods employing LC systems[149, 150]. Despite certain advantages over MALDI-TOF MS, such as absolute quantification and enhanced resolution, the throughput of this approach remains relatively low.

Recently, we have developed a method for the analysis of apo-CIII proteoforms using ultrahigh-resolution MALDI FT-ICR MS[35]. Apo-CIII proteoforms were mainly detected as singly-charged ions. Thus, apo-CIII oxidoforms (1 and 2 times MetO) were detected at +15.995 Th and +31.990 Th from the non-oxidised forms (**Figure 1**). The degree of methionine oxidation of apo-CIII can vary greatly but the complete oxidation of apo-CIII (i.e. 100% conversion of the two methionine residues to MetO) is not commonly observed (**Supplementary Figure S1**). Therefore, for each of the four major proteoforms of apo-CIII, two additional oxidoforms were observed in MALDI FT-ICR MS spectra resulting in twelve proteoforms (**Figure 1**). In addition to that, we were able to detect C-terminal alanine cleaved and fucosylated proteoforms (**Supplementary Table S1**).

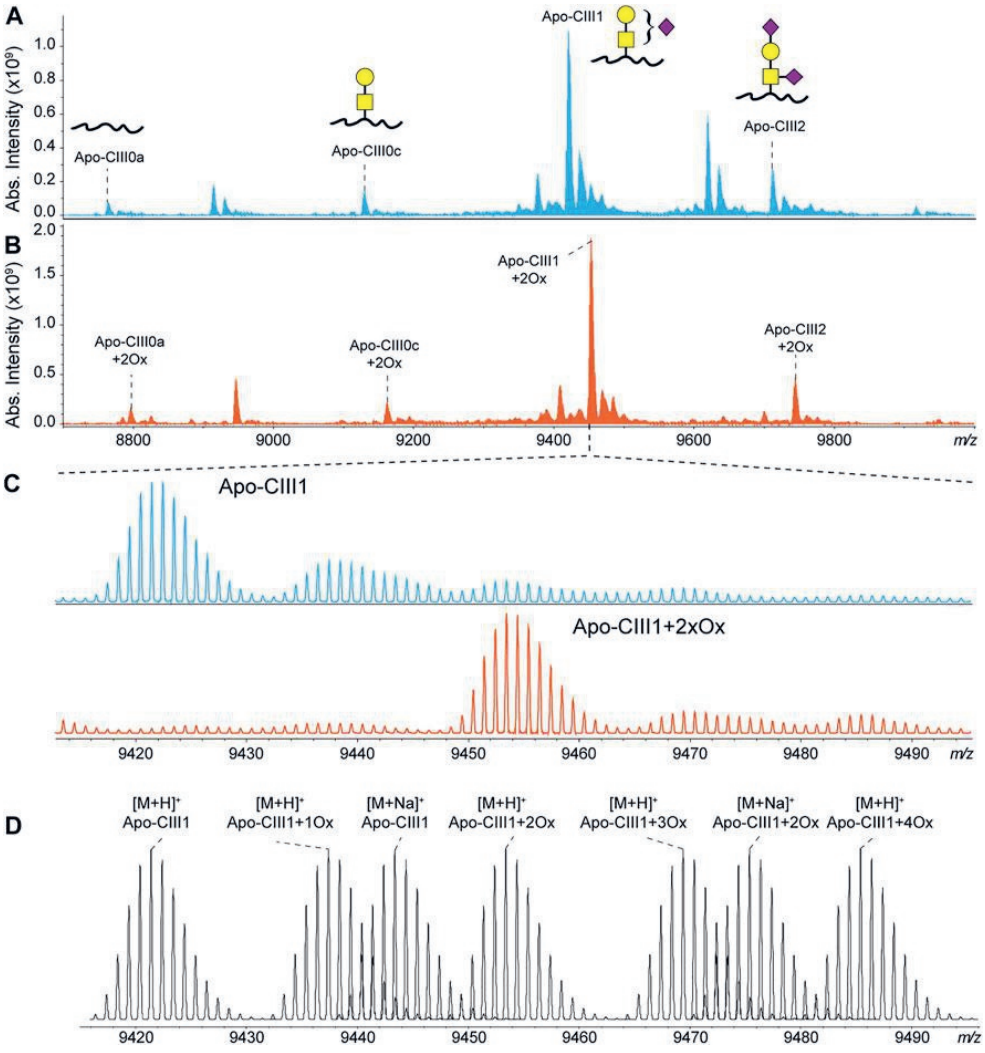


FIGURE 1. MALDI FT-ICR mass spectra of apo-CIII proteoforms from (A) non-treated (blue) and (B) H_2O_2 treated (orange) human blood plasma samples. The four most abundant proteoforms of apo-CIII are annotated. Mass spectra from non-treated samples are characterised by the presence of oxidoforms as exemplified for apo-CIII1 in panel (C). The H_2O_2 treatment allows the almost complete conversion of methionine residues to methionine sulfoxide with a minor conversion to methionine sulfone. Theoretical isotopic distribution of apo-CIII1 proteoforms are depicted in panel (D). For graphical representations of glycan structures: yellow square (*N*-acetylgalactosamine), yellow circle (galactose), purple diamond (*N*-acetylneuraminic acid).

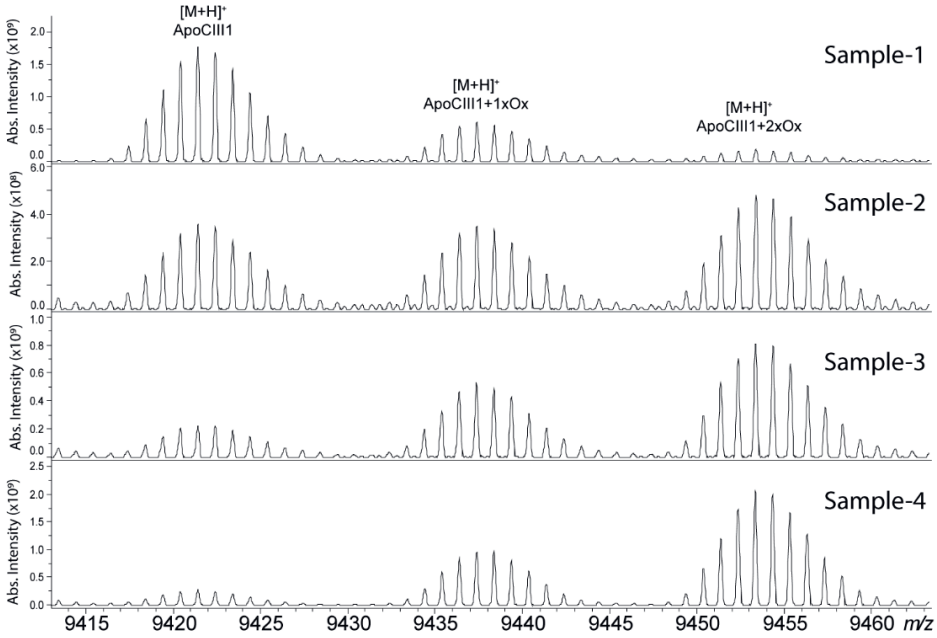


Figure S1. Degree of apo-CIII oxidation in four different plasma samples (without controlled oxidation). The top spectrum was obtained from a fresh aliquot of plasma sample stored at -80 °C after collection. The other spectra were obtained from plasma aliquots that underwent three freeze/thaw cycles.

To reduce sample complexity, we included an oxidation step with hydrogen peroxide to perform a controlled oxidation of both apo-CIII methionine residues to MetO (**Figure 1**). While the implementation of the oxidation step added two hours to the workflow for the analysis, it reduced the heterogeneity of the spectra and facilitated MS data processing using MassyTools software (see Implementation of MassyTools software for high-throughput MS data processing)[140]. The efficiency of the controlled oxidation was tested on 136 standard and 771 clinical plasma samples. The relative intensities between the non-, mono- and di-oxidised forms of apo-$CIII_{0a}$, apo-$CIII_{0c}$, apo-$CIII_1$ and apo-$CIII_2$ are reported in **Supplementary Table S2 and S3**. Oxidation rates over 90% were found for apo-$CIII_1$ and apo-$CIII_2$. Oxidation efficiency seemed to be lower for apo-$CIII_{0a}$, apo-$CIII_{0c,}$ however, close inspection of the spectra revealed the presence of interfering species that contributed to the signal of the non- and mono-oxidised forms of apo-$CIII_{0a}$, apo-$CIII_{0c}$ thus increasing their apparent relative intensity (**Supplementary Figure S2**). Therefore, the controlled oxidation was considered efficient for all four

proteoforms by providing consistent oxidation rates across the standard and clinical plasma samples. These results supported our strategy of using only the signal of the di-oxidised apo-CIII proteoforms for further statistical analysis. The good efficiency and repeatability of the oxidation step allowed us to assess associations between apo-CIII glycosylation and different lipid markers using only the signal of the di-oxidised forms.

Table S1. **Apolipoprotein-CIII proteoforms detected using the MALDI-FT-ICR MS method**. The method allowed to detect four major glycoforms that occur in three different oxido-forms. The combination of glycosylation and oxidation variation results in a total of 12 proteoforms. Detectable C-terminal alanine-cleaved proteoforms are included in the table. TruncA, truncation of a single C-terminal alanine; TruncAA, truncation of two C-terminal alanine residues.

Apolipoprotein-CIII amino acid sequence: SEAEDASLLSFMQGYMKHATKTAKDALSSVQESQVAQ-QARGWVTDGFSSLKDYWSTVKDKFSEFWDLDPEVRPTSAVAA		
Observed apolipoprotein-CIII glycoform	**Calculated *m/z* value for the most abundant peak within an isotopic distribution $[M+H]^+$**	**Ref.**
Apo-$CIII_{0a}$	8765.238	[137]
Apo-$CIII_{0a}$+Ox	8781.233	-
Apo-$CIII_{0a}$+2Ox	8797.228	-
truncA Apo-$CIII_{0a}$+2Ox	8726.191	[137]
truncAA Apo-$CIII_{0a}$+2Ox	8655.154	-
Apo-$CIII_{0C}$	9130.371	[137]
Apo-$CIII_{0C}$+Ox	9146.366	-
Apo-$CIII_{0C}$+2Ox	9162.361	-
truncA Apo-$CIII_{0C}$+2Ox	9091.323	[137]
truncAA Apo-$CIII_{0C}$+2Ox	9019.284	-
Apo-$CIII_1$	9421.466	[137]
Apo-$CIII_1$+Ox	9437.461	-
Apo-$CIII_1$+2Ox	9453.456	-
truncA Apo-$CIII_1$+2Ox	9382.419	[137, 151]
truncAA Apo-$CIII_1$+2Ox	9311.382	-
Apo-$CIII_2$	9712.562	[35, 137]
Apo-$CIII_2$+Ox	9728.557	-
Apo-$CIII_2$+2Ox	9744.552	-
truncA Apo-$CIII_2$+2Ox	9673.514	[151]
truncAA Apo-$CIII_2$+2Ox	9602.477	-

Table S2. Relative intensity (and standard deviation) of the oxidoforms of the most abundant apo-CIII proteoforms in 136 standard plasma samples after controlled oxidation. The high value of the di-oxidised forms indicates a high oxidation efficiency. The relative intensities may be affected by the presence of interfering peaks as shown in Figure S2.

	Non-Ox	Mono-Ox	Di-Ox
apo-$CIII_{0a}$	4.3% (2.0%)	17.0% (5.0%)	78.7% (6.4%)
apo-$CIII_{0c}$	3.3% (0.8%)	11.6% (1.0%)	85.1% (1.7%)
apo-$CIII_{1}$	1.2% (0.4%)	6.6% (0.8%)	92.2% (1.0%)
apo-$CIII_{2}$	0.7% (0.4%)	6.5% (0.9%)	92.8% (1.0%)

Table S3. Relative intensity (and standard deviation) of the oxidoforms of the most abundant apo-CIII proteoforms in 771 control plasma samples after controlled oxidation. The high value of the di-oxidised forms indicates a high oxidation efficiency. The relative intensities of non- and mono-oxidised apo-$CIII_{0a}$ and apo-$CIII_{0c}$ were increased as a result of the presence of interfering species as exemplified in Figure S2.

	Non-Ox	Mono-Ox	Di-Ox
apo-$CIII_{0a}$	8.8% (6%)	22.7% (12.3%)	68.4% (16.3%)
apo-$CIII_{0c}$	4.3% (4.3%)	11.6% (1.8%)	84.1% (5.0%)
apo-$CIII_{1}$	2.0% (0.6%)	7.4% (0.9%)	90.6% (1.2%)
apo-$CIII_{2}$	1.7% (1.1%)	6.3% (1.4%)	91.9% (1.7%)

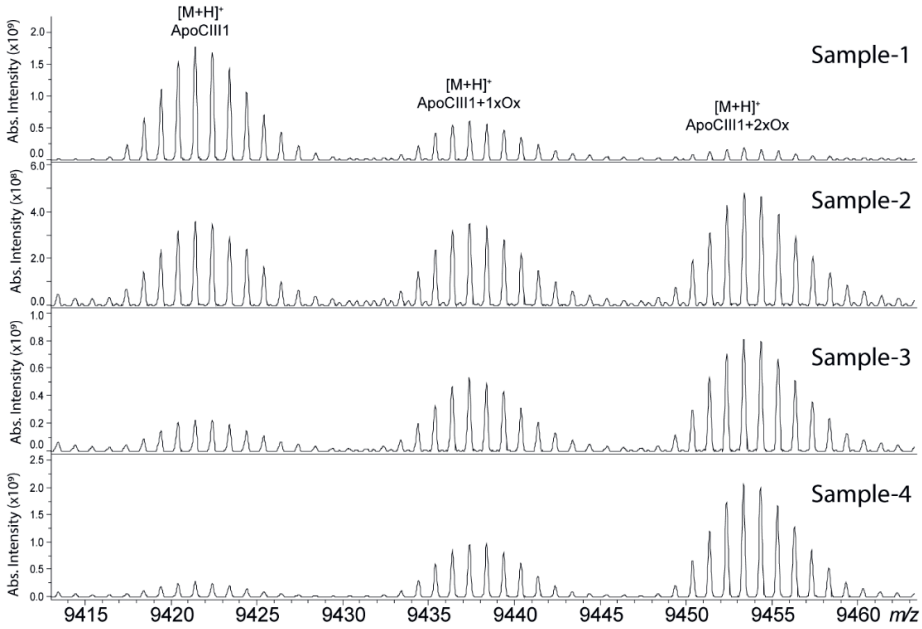


Figure S1. Degree of apo-CIII oxidation in four different plasma samples (without controlled oxidation). The top spectrum was obtained from a fresh aliquot of plasma sample stored at -80 °C after collection. The other spectra were obtained from plasma aliquots that underwent three freeze/thaw cycles.

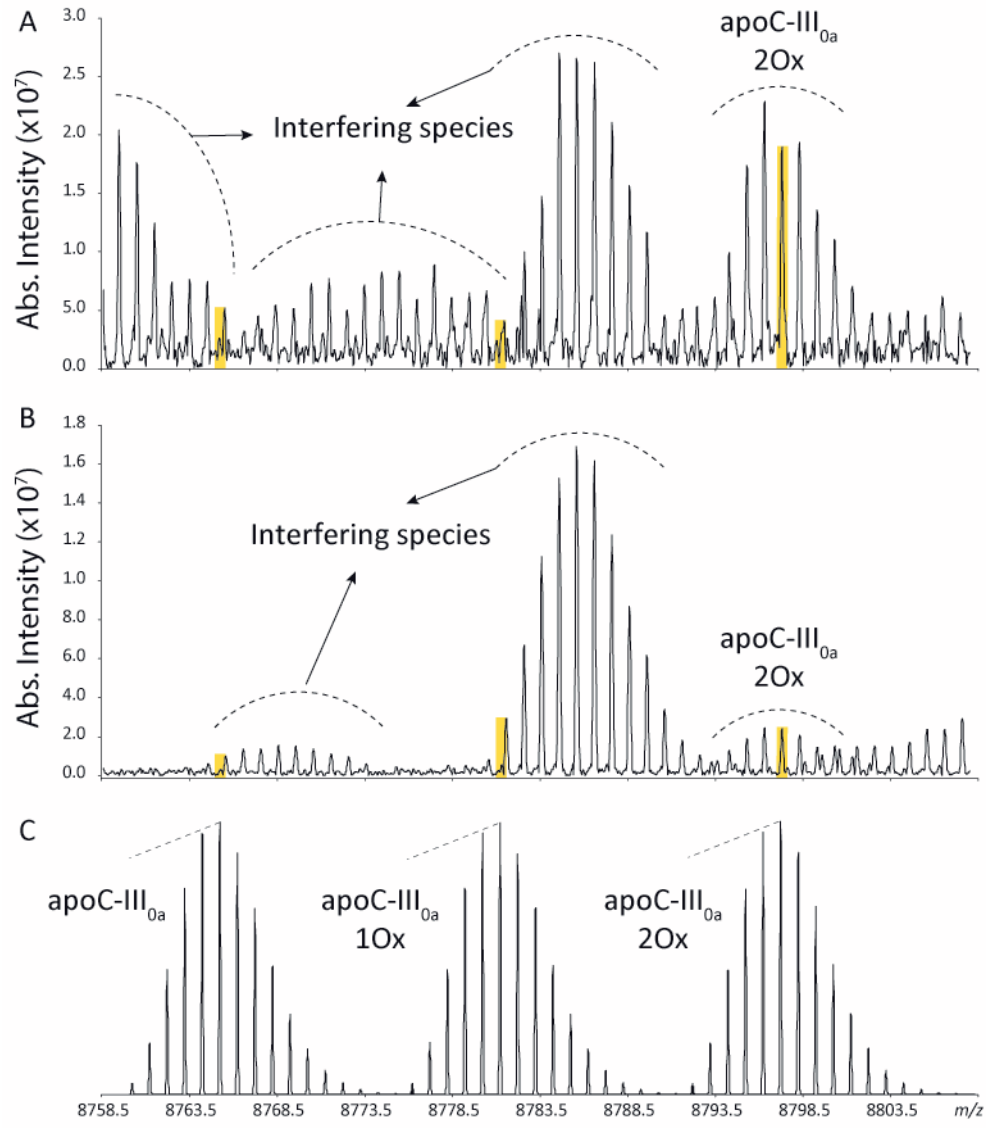


Figure S2. A and B) Enlargements of MALDI FT-ICR MS spectra in the *m/z*-region of apo-CIII$_{0a}$ oxidoforms. The presence of interfering species can affect the determination of the intensity of the most intense isotopic peaks which is calculated in the regions of the spectra highlighted in yellow. In fact, the high relative abundance of Non-Ox and Mono-Ox apo-CIII0a forms (namely 8.8% and 22.7%) in Table S3 was attributed to the presence of interfering species rather than an incomplete oxidation. C) Theoretical isotopic distributions.

Minimizing sialic acid loss using sinapinic acid (SPA) as MALDI matrix

In our previously reported ultrahigh-resolution MALDI FT-ICR MS method for the analysis of apo-CIII proteoforms HCCA was used as a MALDI matrix[35, 137]. This compound was chosen to increase the sensitivity for other serum peptides and small proteins present in C18-SPE eluates obtained from the high-throughput enrichment step using magnetic beads. However, it is known that sialic acid loss can result from in-source decay fragmentation events of glycan structures even when linked to peptides and proteins. In fact, previous reports on the analysis of apo-CIII by MALDI-TOF MS were based on the use of a MALDI matrix colder than HCCA, namely SPA[57-59]. The use of SPA allowed to minimize the loss of sialic acid, as evidenced by an increased relative intensity of both the mono- and the disialylated apo-CIII proteoforms and leading to reproducible apo-CIII glycosylation profiles (**Figure 2 and Supplementary Table S4**). Importantly, compared to other matrices previously used for profiling of apo-CIII glycoforms such as DHB[152, 153], SPA provides more desirable matrix/analyte co-crystallization in the context of high-throughput, automated MALDI measurements.

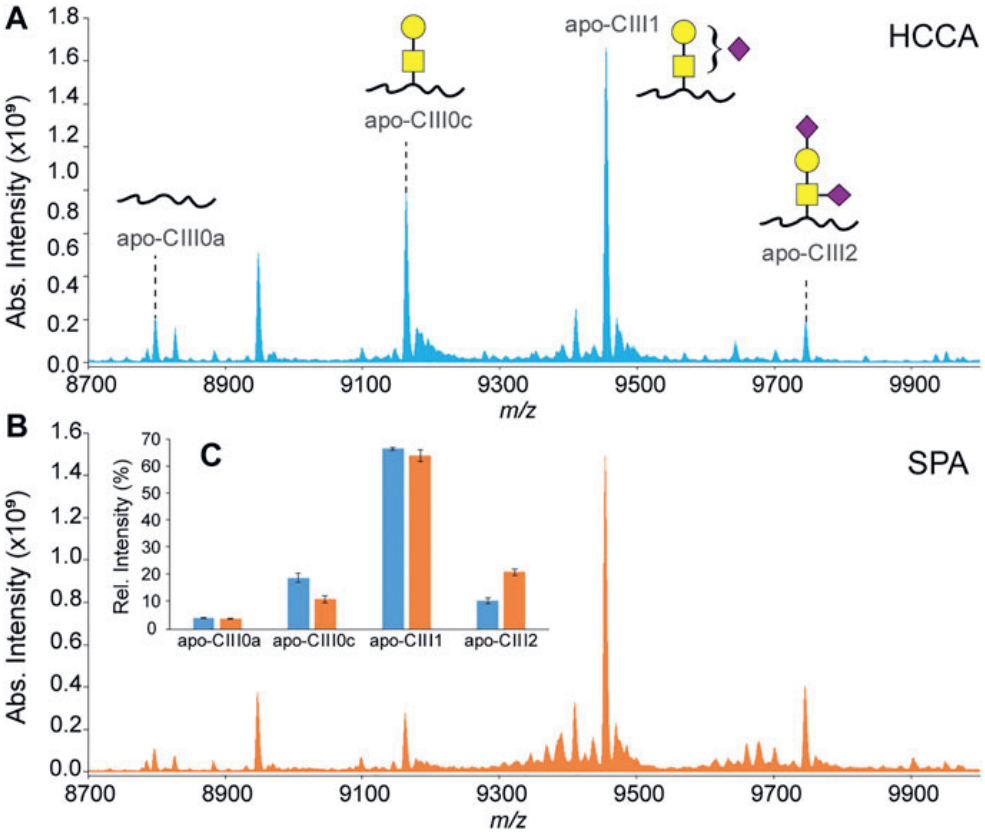


FIGURE 2. MALDI-FT-ICR mass spectra of apo-CIII proteoforms from H_2O_2-treated human blood plasma samples obtained using either HCCA (A) or SPA (B) as MALDI matrices. Relative peak intensities of the four most abundant proteoforms of apo-CIII are shown for both sets (n = 3 in duplicate) of samples in the inset (C).

Table S4. Average relative peak intensities and standard diviation (SD) values obtained from the analysis of three plasma samples in duplicate using HCCA and SPA as a MALDI matrix.

		apo-CIII$_{0a}$	**apo-CIII$_{0c}$**	**apo-CIII$_1$**	**apo-CIII$_2$**
HCCA	Relative Intensity	4%	19%	67%	10%
	SD	0%	2%	1%	1%
SPA	Relative Intensity	4%	11%	64%	21%
	SD	0%	1%	2%	1%

Implementation of MassyTools software for high-throughput MS data processing

One of the advantages of using ultrahigh-resolution MS is that measurements at isotopic resolution provide more spectra information compared to broad-peak detection in linear mode MALDI-TOF MS. Previously, we showed that the goodness of the observed isotopic distributions can be used as a quality control parameter for the selection of high-quality spectra generated from the analysis of a large cohort of samples[154]. This concept was then implemented in a more powerful software – namely, MassyTools – developed for the high-throughput processing of MALDI mass spectra[140]. MassyTools allows the determination of a series of quality control parameters that can be used to perform a curation of MS data at different levels. Mass spectra with unacceptable internal calibration quality and low intensity were discarded at first. Then, the quality of the signal of each apo-CIII proteoform was assessed using the S/N and MME values determined for the most intense peak within an isotopic distribution. Additionally, the quality of such distribution (i.e. IPQ value) was taken into account. The distributions of values of these parameters over 136 standard and 771 clinical plasma samples are reported in **Supplementary Figures S3 and S4**, and **Table S5**. The analytes passing the curation process were then used for statistical analysis.

Table S5. Quality parameters of four major apo-CIII proteoforms: signal to noise (S/N), isotopic pattern quality (IPQ) and mass measurement error (MME; in part-per-million (ppm)). The average values and standard deviations (SD) were calculated based on 136 VisuCon plasma standards and 771 controls used in the cohort study. H_2O_2 treatment provided good quality spectra with analyte peaks meeting minimum two of the set quality criteria: PPM ≤ 10, IPQ ≤ 0.2 and S/N ≥ 9.

			Apo-$CIII_{0a}$+2Ox *m/z* = 8797.228	**Apo-$CIII_{0c}$+2Ox *m/z* = 9162.361**	**Apo-$CIII_1$+2Ox *m/z* = 9453.456**	**Apo-$CIII_2$+2Ox *m/z* = 9744.552**
Plasma standards (n = 136)	**S/N**	**Mean**	**38.37**	**70.29**	**127.48**	**62.42**
		SD	17.80	21.45	27.69	17.11
	IPQ	**Mean**	**0.03**	**0.02**	**0.01**	**0.02**
		SD	0.02	0.01	0.01	0.01
	MME (ppm)	**Mean**	**0.65**	**1.31**	**1.41**	**0.48**
		SD	1.66	0.37	0.39	0.13
Controls (n = 771)	**S/N**	**Mean**	**22.03**	**53.59**	**84.45**	**44.86**
		SD	15.18	17.96	17.13	10.84
	IPQ	**Mean**	**0.09**	**0.04**	**0.02**	**0.03**
		SD	0.09	0.03	0.01	0.03
	MME (ppm)	**Mean**	**1.64**	**0.98**	**1.04**	**0.37**
		SD	4.36	0.6	0.5	0.46

As assessed on 136 standard plasma samples, which were distributed over 17 MALDI target plates measured over 28 days, the method provided good repeatability for relative quantitation of all four proteoforms with CVs in a range of 1% to 18% for average intra-plate and 6% to 16% for inter-plate variability (**Table 1**). While we reduced in-source decay by selecting SPA as a MALDI matrix, we expect that partial sialic acid loss from apo-$CIII_1$ during MS analysis may lead to a slight, artificial increase in the apo-$CIII_{0c}$ glycoform abundance. Hence, fluctuations in the extent of sialic acid loss may contribute to the larger CVs for apo-$CIII_{0c}$.

Table 1. Inter- and intra-plate variability. Relative peak intensities, standard deviation (SD) and coefficient of variation (CV) are given for the intra- and inter-plate variability based on 136 plasma standards.

		Apo-$CIII_{0a}$ +2Ox	**Apo-$CIII_{0c}$ +2Ox**	**Apo-$CIII_1$ +2Ox**	**Apo-$CIII_2$ +2Ox**
Inter-plate	Relative peak intensity	0.039	0.114	0.636	0.203
	SD	0.003	0.018	0.04	0.02
	CV	8%	16%	6%	10%
Average intra-plate	CV	7%	18%	1%	10%

Associations between apo-CIII sialylation and lipid markers

We used this approach to determine non-glycosylated and the glycosylated non-sialylated, mono-sialylated and disialylated apo-CIII glycoforms within a cohort of 746 individuals without diabetes (cohort characteristics in **Table 2**) and test their association with a range of metabolic biomarkers. We found the association of disialylated apo-$CIII_2$ with overall improved lipid profiles and decreased BMI (**Table 3 and Supplementary Table S6**), which is in accordance with some of the previous reports[57, 59]. A subgroup analysis in participants not using statins or fibrates, did not change these associations (**Supplementary Table S7A and 7B**).

Table 2. DiaGene cohort characteristics.

	Individuals without diabetes
Participants, n	746
Male sex, n (%)	290 (38.9)
Age, year	65.7 (6.7)
Age within males, year	66.0 (6.7)
Age within females, year	65.5 (6.7)
BMI, kg/m2	25.4 (4.5)
HDL-cholesterol, mmol/l	1.48 (0.36)
non-HDL-cholesterol, mmol/l	4.09 (0.97)
LDL-cholesterol, mmol/l	3.56 (0.90)
Triglycerides, mmol/l	1.20 (0.68)
Total cholesterol, mmol/l	5.57 (0.99)
Use of lipid lowering therapy, n (%)	93 (12.5)

Mean (and standard deviation) for normal distribution and median (and interquartile range) for non-normal distributions (BMI and triglycerides).

Table S6. Associations of apo-CIII sialylation with clinical characteristics. The sum of the glycoforms Apo-CIII0c, Apo-CIII1, and Apo-CIII2 was set to 1.0.

Characteristics	Apo-$CIII_{0c}$			Apo-$CIII_1$			Apo-$CIII_2$		
	Male	Female	p-value	Male	Female	p-value	Male	Female	p-value
Sex (mean ± SD)	0.114 ± 0.026	0.116 ± 0.025	0.531	0.667 ± 0.028	0.674 ± 0.031	**0.001**	0.218 ± 0.036	0.209 ± 0.039	**0.003**
	beta		p-value	beta		p-value	beta		p-value
Age	-9.22E-05		5.05E-01	-2.96E-04		7.06E-02	4.06E-04		5.11E-02
BMI	8.34E-04		**8.15E-04**	1.55E-03		**1.16E-07**	-2.42E-03		**6.13E-11**
HDL cholesterol	-1.09E-03		6.80E-01	-6.35E-03		**4.21E-02**	7.66E-03		5.30E-02
non-HDL cholesterol	1.39E-03		1.64E-01	7.60E-03		**6.54E-11**	-9.00E-03		**1.11E-09**
LDL cholesterol	1.66E-03		1.20E-01	7.55E-03		**1.59E-09**	-9.23E-03		**6.45E-09**
Triglycerides	1.92E-03		1.01E-01	1.50E-02		**1.99E-20**	-1.69E-02		**8.31E-17**
Total cholesterol	1.18E-03		2.26E-01	6.42E-03		**2.01E-08**	-7.59E-03		**1.75E-07**

Blue: negative associations, red: positive associations, bold: significant p-value. P-values of logarithmically transformed triglyceride concentrations, beta of non-transformed concentrations.

Table S7a. Subgroup analysis in non-diabetic controls, not using statins or fibrates. The sum of the glycoforms Apo-CIII0a, Apo-CIII0c, Apo-CIII1, and Apo-CIII2 was set to 1.0.

	Apo-$CIII_{0a}$		Apo-$CIII_{0c}$		Apo-$CIII_1$		Apo-$CIII_2$	
	beta	p-value	beta	p-value	beta	p-value	beta	p-value
Age	2.37E-05	8.44E-01	-1.18E-04	3.84E-01	-2.63E-04	1.02E-01	3.60E-04	1.04E-01
BMI	-5.76E-04	**8.24E-03**	6.48E-04	**8.19E-03**	1.82E-03	**1.86E-10**	-1.91E-03	**1.49E-06**
HDL cholesterol	1.46E-03	5.26E-01	5.45E-05	9.83E-01	-9.24E-03	**2.20E-03**	7.67E-03	6.56E-02
non-HDL cholesterol	9.09E-04	3.19E-01	1.86E-03	6.75E-02	6.23E-03	**1.55E-07**	-8.85E-03	**6.12E-08**
LDL cholesterol	1.08E-03	2.72E-01	2.51E-03	**2.25E-02**	6.14E-03	**1.80E-06**	-9.56E-03	**6.63E-08**
Triglycerides	2.12E-03	**4.65E-02**	7.37E-04	4.89E-01	1.19E-02	**7.20E-14**	-1.46E-02	**4.68E-12**
Total cholesterol	1.13E-03	2.14E-01	1.86E-03	6.72E-02	4.74E-03	**6.65E-05**	-7.59E-03	**3.40E-06**

Blue: negative associations, red: positive associations, bold: significant p-value. P-values of logarithmically transformed triglyceride concentrations, beta of non-transformed concentrations.

Table S7b. Subgroup analysis in non-diabetic controls, not using statins or fibrates. The sum of the glycoforms Apo-CIII0c, Apo-CIII1, and Apo-CIII2 was set to 1.0.

	Apo-$CIII_{0c}$		Apo-$CIII_1$		Apo-$CIII_2$	
	beta	p-value	beta	p-value	Beta	p-value
Age	-1.20E-04	4.07E-01	-2.49E-04	1.51E-01	3.81E-04	8.52E-02
BMI	5.97E-04	**2.25E-02**	1.47E-03	**1.96E-06**	-2.13E-03	**7.33E-08**
HDL cholesterol	1.96E-04	9.43E-01	-8.52E-03	**8.87E-03**	8.50E-03	**4.10E-02**
non-HDL cholesterol	1.99E-03	6.74E-02	7.04E-03	**3.80E-08**	-9.05E-03	**2.96E-08**
LDL cholesterol	2.69E-03	**2.19E-02**	6.99E-03	**4.41E-07**	-9.73E-03	**3.67E-08**
Triglycerides	9.42E-04	4.00E-01	1.41E-02	**4.21E-17**	-1.49E-02	**2.14E-12**
Total cholesterol	2.01E-03	6.43E-02	5.66E-03	**1.01E-05**	-7.66E-03	**2.73E-06**

Blue: negative associations, red: positive associations, bold: significant p-value. P-values of logarithmically transformed triglyceride concentrations, beta of non-transformed concentrations.

So far, the inhibitory effect of apo-CIII on LPL has been linked to total apo-CIII concentration, but not to the relative proportion of apo-CIII glycoforms[150]. Recent studies proposed that the presence of apo-CIII on triglyceride-rich lipoproteins (TRLs) alters the affinity between TRLs and their receptors in the liver. Kegulian *et al.* demonstrated that the degree of apo-CIII sialylation directs TRLs to different hepatic clearance pathways, as shown in mice[59]. In detail, apo-$CIII_1$-enriched very low-density lipoproteins (VLDLs) are preferentially cleared by faster-acting low-density lipoprotein (LDL) receptor (LDLR) and LDL receptor-related protein 1 (LRP1), whereas apo-$CIII_2$

directs VLDLs to syndecan 1 (SDC1) receptors that are characterised by a slower but larger capacity metabolism of TRLs. The same study also showed that a 13-week antisense oligonucleotide treatment for apo-CIII, which, as expected, reduced plasma TG levels, also altered relative abundances of these two glycoforms leading to an increase of apo-$CIII_2$ and a decrease of apo-$CIII_1$. The increase of the apo-$CIII_2$/apo-$CIII_1$ ratio in a response to the antisense oligonucleotide therapy was explained by a differing capacity and clearance speed of the hepatic TRL receptors. In support of this, we observed in our cohort study of individuals without diabetes a positive association of the relative abundance of apo-$CIII_1$ glycoform with TG levels, and a negative association for apo-$CIII_2$ (**Table 3 and Figure 3**).

Table 3. Associations of apo-CIII glycosylation with clinical characteristics.

	Apo-CIII$_{0a}$			Apo-CIII$_{0c}$			Apo-CIII$_1$			Apo-CIII$_2$		
Seks	Male	Female	p-value	Male	Female	p-value	Male	Female	p-value	Male	Female	p-value
(mean ± SD)	0.05 ± 0.02	0.05 ± 0.02	0.605	0.11 ± 0.02	0.11 ± 0.02	0.558	0.63 ± 0.03	0.64 ± 0.03	**0.002**	0.21 ± 0.04	0.20 ± 0.04	**0.004**
	beta	p-value		beta	p-value		beta	p-value		beta	p-value	
Age	1.28E-05	9.10E-01		-8.84E-05	4.96E-01		-2.97E-04	5.12E-02		3.84E-04	6.47E-02	
BMI	-4.61E-04	**2.23E-02**		8.60E-04	**2.31E-04**		1.81E-03	**1.74E-11**		-2.21E-03	**2.41E-09**	
HDL cholesterol	1.52E-03	4.80E-01		-1.20E-03	6.29E-01		-7.23E-03	**1.24E-02**		6.87E-03	8.27E-02	
non-HDL cholesterol	1.36E-03	9.56E-02		1.21E-03	1.95E-01		6.53E-03	**1.46E-09**		-8.88E-03	**1.81E-09**	
LDL cholesterol	1.55E-03	7.55E-02		1.45E-03	1.49E-01		6.41E-03	**3.32E-08**		-9.15E-03	**8.51E-09**	
Triglycerides	2.61E-03	**7.45E-03**		1.60E-03	1.53E-01		1.25E-02	**9.02E-16**		-1.66E-02	**1.44E-16**	
Total cholesterol	1.51E-03	5.82E-02		9.98E-04	2.75E-01		5.28E-03	**6.77E-07**		-7.58E-03	**1.76E-07**	

Blue: negative associations, red: positive associations, bold: significant p-value. P-values of logarithmically transformed triglyceride concentrations, beta of non-transformed concentrations.

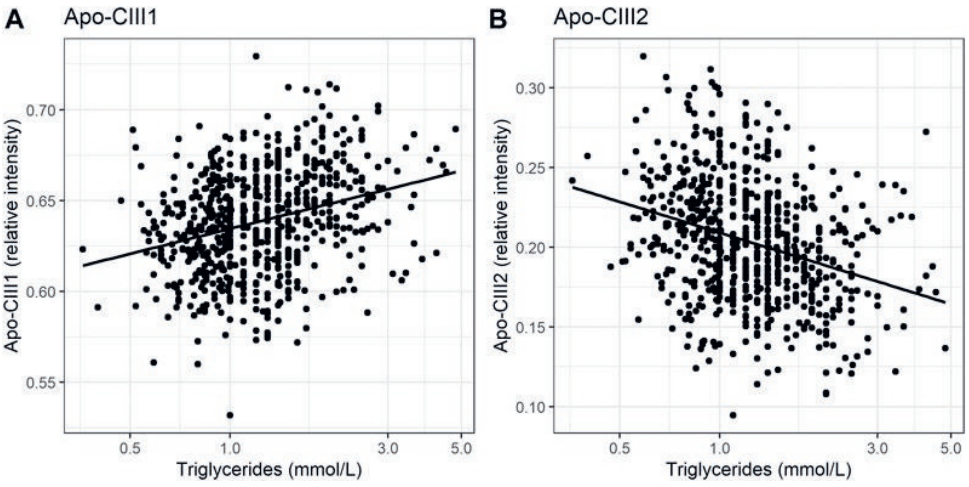


Figure 3. Relationship of apo-CIII1 (A) and apo-CIII2 (B) with triglycerides.

Our study supports the hypothesis that triglyceride clearance may be regulated, or at least strongly influenced, by apo-CIII glycosylation, specifically sialylation. However, other aspects have to be considered. For instance, defective LDLR/LRP1-driven metabolic pathways might lead to decreased clearance of TGs. Expression and stability of LDLR and LRP1 in the liver might be affected by naturally occurring genetic

variants[132, 155-157]. Moreover, it has been shown in mice that a high-fat diet can lead to the down-regulated expression of hepatic LRP1 by causing hyperglycemia with a high level of plasma triglycerides[158]. In humans, obesity is associated with increases in plasma triglycerides[159]. We hypothesised that the association of apo-CIII glycoforms with triglycerides could be confounded by BMI. Surprisingly, after adjustment for BMI, the direction of effect and goodness-of-fit did not evidently change (**Supplementary Table S8a and 8b**). This indicates that the association of apo-CIII sialylation with triglycerides is largely independent of BMI, and that it is not obesity-associated physiological changes that determine apo-CIII sialylation and its association with triglycerides.

Expression levels of apo-CIII were not investigated in this study. The differences in apo-CIII glycosylation profiles observed between individuals may be caused by varying expression levels of apo-CIII[150] or apo-CIII glycoforms[135], or the accumulation of certain glycoforms due to dysfunctional clearance pathways, based on recent findings by Kegulian et al.[59]. It may also be a combination of the listed factors, which should be explored in further research. Nevertheless, from the results of this study, we cannot determine whether apo-CIII sialylation influences triglyceride levels, or vice versa. Further studies are needed to elucidate the genetic and environmental factors that determine apo-CIII sialylation in health and disease.

Table S8a. Associations of apo-CIII glycoforms with triglycerides, with and without adjustment for BMI. The sum of the glycoforms Apo-CIII0a, Apo-CIII0c, Apo-CIII1, and Apo-CIII2 was set to 1.0.

	Apo-$CIII_{0a}$			Apo-$CIII_{0c}$			Apo-$CIII_1$			Apo-$CIII_2$		
	beta	p-value	R2	beta	p-value	R2	beta	p-value	R2	beta	p-value	R2
Triglycerides	2.61E-03	**7.45E-03**	0.01	1.60E-03	1.53E-01	0.00	1.25E-02	**9.02E-16**	0.09	-1.66E-02	**1.44E-16**	0.09
Triglycerides adjusted for BMI	3.69E-03	**3.64E-04**	0.02	1.66E-05	7.96E-01	0.02	1.05E-02	**3.70E-11**	0.11	-1.43E-02	**2.72E-12**	0.09

Blue: negative associations, red: positive associations, bold: significant p-value. P-value and R-square (R2) are calculated for the logarithmically transformed values of triglycerides. 'Triglycerides adjusted for BMI' represent a multivariate linear model, BMI and triglycerides as independent variables.

Table S8b. Associations of apo-CIII glycoforms with triglycerides, with and without adjustment for BMI. The sum of the glycoforms Apo-CIII0c, Apo-CIII1, and Apo-CIII2 was set to 1.0.

	Apo-$CIII_{0c}$			Apo-$CIII_1$			Apo-$CIII_2$		
	beta	p-value	R2	beta	p-value	R2	beta	p-value	R2
Triglycerides	1.92E-03	1.01E-01	0.00	1.50E-02	**1.99E-20**	0.12	-1.69E-02	**8.31E-17**	0.10
Triglycerides adjusted for BMI	4.12E-04	5.54E-01	0.01	1.38E-02	**1.10E-16**	0.12	-1.42E-02	**4.78E-12**	0.11

Blue: negative associations, red: positive associations, bold: significant p-value. P-value and R-square (R2) are calculated for the logarithmically transformed values of triglycerides. 'Triglycerides adjusted for BMI' represent a multivariate linear model, BMI and triglycerides as independent variables.

CONCLUSIONS

Apo-CIII is a novel potential drug target in the management of cardiovascular disease driven by multiple studies demonstrating that plasma levels of apo-CIII are predictive of coronary heart disease and the risk of disease-related events[160]. Previously, it has been shown that sialylated apo-CIII glycoforms are differentially cleared by hepatic receptors and that a higher apo-$CIII_2$/apo-$CIII_1$ ratio is associated with improved triglyceride levels[59]. In humans, the production rates of these two glycoforms are comparable[161], therefore varying apo-CIII2/apo-CIII1 ratios between individuals in healthy and disease groups might suggest various dysfunctional mechanisms involved in their production and clearance. This is the first large-scale study of apo-CIII glycosylation. Clinical cohort studies employing large numbers of individuals will provide more insight into this topic, and the development of highly robust and accurate analytical methods enabling such large-scale studies is warranted.

Here, we present a workflow for high-throughput MALDI-FT-ICR MS analysis of apo-CIII glycosylation in human plasma samples varying in protein oxidation levels. The controlled oxidation of apo-CIII methionine residues, the use of sinapinic acid as a MALDI matrix, and the use of MassyTools software for semi-automated, standardised spectra processing have been implemented to achieve highly repeatable measurements of intact apo-CIII proteoforms. The new analytical workflow allowed us to overcome the problem of the high spectral heterogeneity produced by methionine oxidation thus allowing the robust screening of a large cohort of plasma samples for the relative quantitation of apo-CIII proteoforms. Importantly, the evaluation of MS spectra-derived quality parameters was implemented to minimize biases and ensure accuracy of collected data.

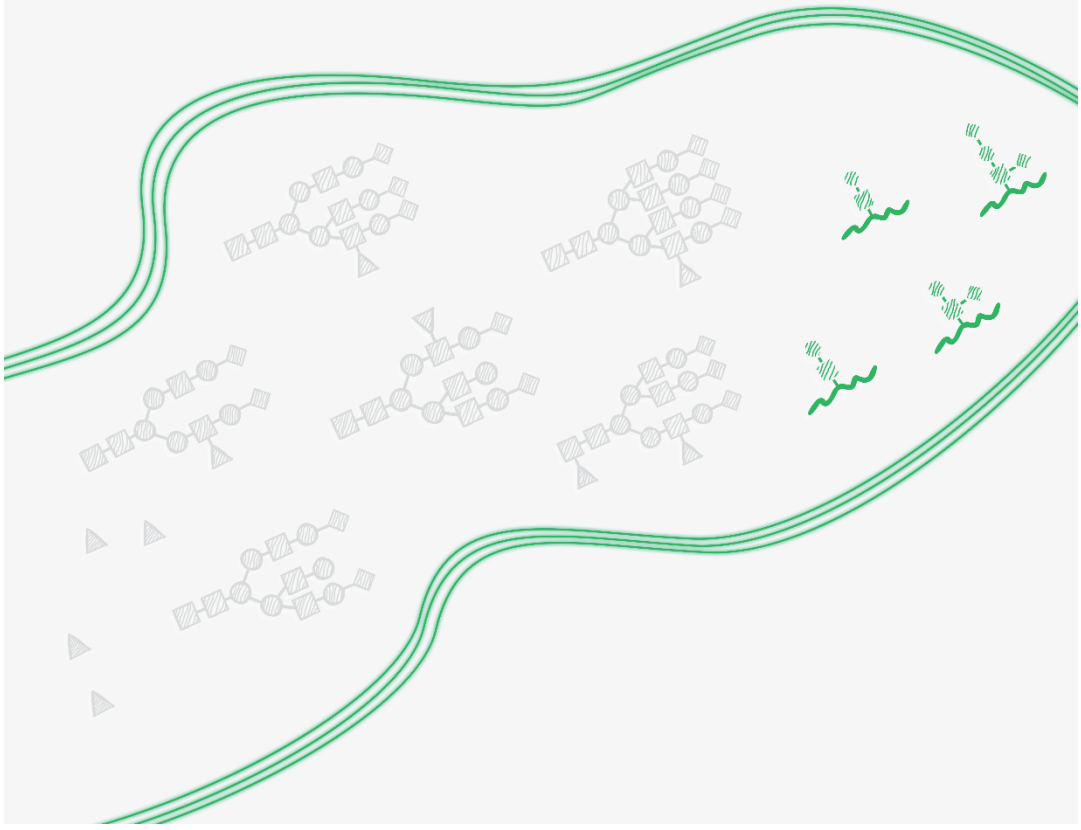
The cohort analysis confirmed that the level of apo-CIII sialylation is strongly associated with lipid markers, especially with triglyceride levels. The relation between relative abundances of apo-CIII glycoforms and cardiovascular disease development should be further explored. More insight into the role of apo-CIII glycosylation in disease

pathophysiology could provide new drug targets. Also, understanding of the mechanisms of existing drugs might increase by considering apo-CIII glycosylation. The methods presented, will enable such large-scale studies.

Acknowledgments

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 722095.

CHAPTER 5



5. Apolipoprotein CIII *O*-glycosylation, the missing link between GALNT2 and plasma lipids.

Annemieke Naber[1], Daniel Demus[2], Roderick Slieker[3,4], Simone Nicolardi[2], Joline W.J. Beulens[4,5,6], Petra J.M. Elders[7], Aloysius G. Lieverse[8], Eric J.G. Sijbrands[1], Leen M. 't Hart[3,4,5,9], Manfred Wuhrer[2] and Mandy van Hoek[1]

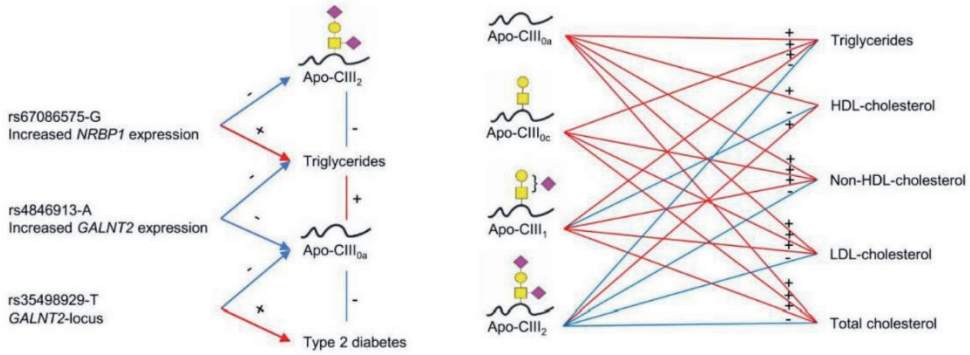
[1] Department of Internal Medicine, Erasmus MC University Medical Center Rotterdam, PO-box 2040, 3000 CA Rotterdam, The Netherlands
[2] Center for Proteomics and Metabolomics, Leiden University Medical Center, PO-box 9600, 2300 RC Leiden, The Netherlands
[3] Department of Cell and Chemical Biology, Leiden University Medical Center, PO-box 9600, 2300 RC Leiden, The Netherlands
[4] Department of Epidemiology and Data Science, Amsterdam UMC, location Vrije Universiteit Amsterdam, Postbus 7057, 1007 MB Amsterdam, The Netherlands
[5] Amsterdam Public Health, Amsterdam Cardiovascular Sciences, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands
[6] Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Postbus 85500, 3508 GA Utrecht, The Netherlands
[7] Department of General Practice, Amsterdam Public Health Institute, Amsterdam UMC, location VUmc, Postbus 7057, 1007 MB Amsterdam, The Netherlands
[8] Department of Internal Medicine, Maxima Medical Center, Postbus 90052, 5600 PD Eindhoven, The Netherlands
[9] Department of Biomedical Data Science, section Molecular Epidemiology, Leiden University Medical Center, Postal zone S5-P PO Box 9600, 2300 RC Leiden, The Netherlands

Reprinted and adapted: Naber, A.; Demus, D.; Slieker, R.; Nicolardi, S.; Beulens, J.W.J.; Elders, P.J.M.; Lieverse, A.G.; Sijbrands, E.J.G.; 't Hart, L.M.; Wuhrer, M.; et al. Apolipoprotein-CIII O-Glycosylation, a Link between GALNT2 and Plasma Lipids. Int. J. Mol. Sci. 2023, 24, 14844. https://doi.org/10.3390/ijms241914844

ABSTRACT

Apolipoprotein-CIII is involved in triglyceride-rich lipoprotein metabolism and linked to beta-cell damage, insulin resistance, and cardiovascular disease. Apolipoprotein-CIII exists in four main proteoforms: non-glycosylated (apo-$CIII_{0a}$), and glycosylated apolipoprotein-CIII with zero, one or two sialic acids (apo-$CIII_{0c}$, apo-$CIII_1$ and apo-$CIII_2$). Our objective is to determine how apo-CIII glycosylation affects lipid traits and type 2 diabetes prevalence, and to investigate the genetic basis of these relations with a GWAS on apo-CIII glycosylation**.** We conducted GWAS on the four apolipoprotein-CIII proteoforms in the DiaGene study in people with and without type 2 diabetes (n=2,318). We investigated the relations of the identified genetic loci and apolipoprotein-CIII glycosylation with lipids and type 2 diabetes. The associations of the genetic variants with lipids were replicated in the Diabetes Care System (n=5,409). Rs4846913-A, in the *GALNT2*-gene, was associated with decreased apo-$CIII_{0a}$. This variant was associated with increased high-density lipoprotein cholesterol and decreased triglycerides, while high apo-$CIII_{0a}$ was associated with raised HDL-cholesterol and triglycerides. Rs67086575-G, located in the *IFT172*-gene, was associated with decreased apo-$CIII_2$ and with hypertriglyceridemia. In line, apo-$CIII_2$ was associated with low triglycerides. On a genome-wide scale, we confirmed that the *GALNT2*-gene plays a major role in *O*-glycosylation of apolipoprotein-CIII, with subsequent associations with lipid parameters. We newly identified the *IFT172*/*NRBP1* region, in literature previously associated with hypertriglyceridemia, as involved in apolipoprotein-CIII sialylation and hypertriglyceridemia. These results link genomics, glycosylation and lipid metabolism, and represent a key step towards unravelling the importance of *O*-glycosylation in health and disease.

GRAPHICAL ABSTRACT



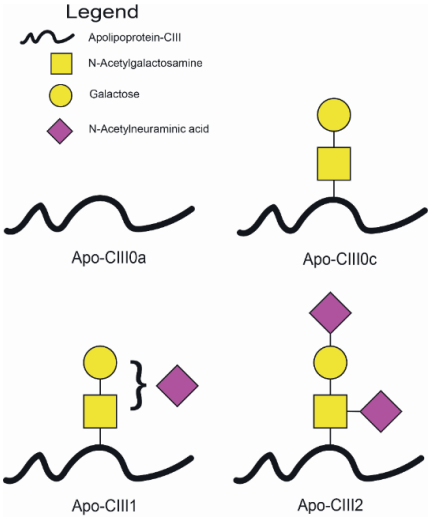


Figure S1. The four most abundant apo-CIII proteoforms. For graphical representations of glycan structures: yellow square (*N*-acetylgalactosamine), yellow circle (galactose), purple diamond (*N*-acetylneuraminic acid).

INTRODUCTION

Type 2 diabetes and cardiovascular disease are two of the leading causes of disease burden worldwide[162]. Triglycerides and low-density lipoprotein (LDL)-cholesterol (c) are important contributors to cardiovascular disease risk[163]. One of the central players in lipid metabolism is apolipoprotein-CIII (apo-CIII)[160].

Apo-CIII inhibits lipoprotein lipase[164] and hepatic lipase[165], which regulate triglyceride-rich lipoprotein (TRL) metabolism[166]. Apo-CIII is present mostly on TRLs when the triglyceride levels are high and mostly on high-density lipoproteins (HDL) when the triglyceride levels are low[167, 168]. Apo-CIII levels and variation in the *APOC3* gene contribute to dyslipidaemia, artery wall inflammation, atherogenesis, and cardiovascular disease (CVD) risk[169-171]. Moreover, high apo-CIII levels have been linked to insulin resistance and beta-cell destruction[172, 173]. *APOC3* promotor variants increase type 2 diabetes risk and the need for insulin treatment, particularly among lean patients[174].

Polypeptide N-acetylgalactosaminyltransferase 2 (GalNAc-T2), encoded by the *GALNT2* gene, initiates mucin-type *O*-glycosylation of apo-CIII in the Golgi apparatus[175]. Mutations in the *GALNT2* gene cause rare congenital glycosylation disorders affecting apo-CIII glycosylation[153] and affect lipid profiles[176]. Without such mutations, apo-CIII exists in different proteoforms[35, 58]. The four most abundant apo-CIII proteoforms are the non-glycosylated apo-$CIII_{0a}$, and the proteoforms containing a mucin-type core-1 *O*-glycosylation with zero (apo-$CIII_{0c}$), one (apo-$CIII_1$), and two (apo-$CIII_2$) sialic acids (**Supplementary Figure S1**)[35, 177]. Apo-CIII proteoforms have different effects on hepatic TRL clearance receptors and on the potential of apo-CIII to inhibit lipoprotein lipase (LPL) resulting in variation of hepatic triglyceride clearance and lipid profile[59, 178]. In smaller scale cohorts, the concentrations of apo-$CIII_{0a}$, apo-$CIII_{0b}$ and apo-$CIII_1$ have been associated with lower small dense LDL[218] and higher triglyceride levels[57, 58, 179].

The objective of our study was to investigate the association between apo-CIII glycosylation and lipid traits, as well as the prevalence of type 2 diabetes, in a large population-based cohort. Moreover, we conducted the first genome-wide association

studies (GWASs) of apo-CIII *O*-glycosylation in the DiaGene Study and explored the associations between genetic variants, apoCIII proteoforms, lipid levels, and type 2 diabetes. The genetic associations with lipid levels were subsequently replicated in the Hoorn Diabetes Care System cohort.

MATERIALS AND METHODS

DiaGene Study

We used cross-sectional data from the DiaGene Study. This study has been described in more detail elsewhere[141]. Briefly, this case-control cohort comprises type 2 diabetes patients from all lines of care and people without diabetes from the area of Eindhoven, the Netherlands. We had available data on apo-CIII *O*-glycosylation for 1,572 persons with type 2 diabetes and 746 persons without diabetes, of which for 1,872 participants also genetic data were available to perform the GWAS analysis.

The Hoorn Diabetes Care System cohort

For replication of the genetic associations with lipid levels, we used data from the Hoorn Diabetes Care System (Hoorn DCS) cohort, described in detail elsewhere[180]. Patients with type 2 diabetes treated in primary care in the region of West Friesland, the Netherlands, are included in this cohort. Genetic, biochemical, anthropometric and clinical data of 5,409 persons with type 2 diabetes were available. Apo-CIII glycosylation measurements were not available in the Hoorn DCS cohort.

Informed consent was obtained from all subjects involved in the study. Both studies were approved by the Medical Ethics Committees of the involved hospitals in compliance with the Declaration of Helsinki principles (DiaGene MEC-2004-230, Hoorn DCS 2007/57).

Apo-CIII glycosylation measurements

The analysis of apo-CIII *O*-glycosylation in the DiaGene Study was performed using a high-throughput method based on solid-phase extraction, matrix-assisted laser desorption/ionization (MALDI) and ultrahigh-resolution Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometry (MS), described by Demus et al.[179]. Apo-$CIII_{0a}$ represents non-glycosylated apo-CIII; the non-, mono- and di-sialylated glycoforms are described as apo-$CIII_{0c}$, apo-$CIII_1$ and apo-$CIII_2$, respectively. Apo-$CIII_{0a}$ was normalised to the sum of all four proteoforms: apo-$CIII_{0a}$, apo-$CIII_{0c}$, apo-$CIII_1$ and apo-$CIII_2$ to reflect the proportion of non-glycosylated apo-CIII. The sum of apo-$CIII_{0c}$, apo-$CIII_1$, and apo-$CIII_2$ was set to 1.0 to obtain the proportion of these three glycoforms within all glycosylated species of apo-CIII. Additionally, the ratio of apo-$CIII_1$ and apo-$CIII_2$ was

calculated. Individuals with missing glycosylation data, missing age, sex, or genetic principal components were excluded from the analyses. Glycan measurements were natural log-transformed prior to GWAS analysis, because of the skewness of their distributions.

GWAS of apo-CIII glycosylation

In the DiaGene and Hoorn DCS cohorts, quality control was performed in PLINK[181], and genotypes were imputed to the Haplotype Reference Consortium r1.1 reference panel[182, 183] using the Michigan Imputation Server[184]. Variants with an imputation quality <0.4, minor allele frequency ≤0.05 or ≥0.95; effective allele count ≤5; or when out of Hardy-Weinberg Equilibrium (HWE) with HWE p-value $\leq 1*10^{-4}$ were excluded. Using zCall[185] we aimed to call previously uncalled genotypes, mostly of rare variants. Manhattan plots, QQ plots and locus zoom plots were generated using FUMA[186]. In the DiaGene study, we conducted separate GWASs on all four apo-CIII proteoforms, and the apo-$CIII_1$/apo-$CIII_2$ ratio with RVtest[187] and quality control was performed with EasyQC[188]. We assessed genetic associations using linear regression models adjusted for age, sex and significant principal components of ancestry (**Supplementary Table 1**). P-values lower than or equal to $5*10^{-8}$ were considered genome-wide significant. To differentiate diabetes-specific and general genetic effects, the associations of the genetic variants with the apo-CIII proteoforms were investigated separately for people with or without type 2 diabetes, using linear regression models in SPSS.

Table S1. Association of apolipoprotein-CIII proteoforms with principal components of ancestry.

	C1 p-value	C2 p-value	C3 p-value	C4 p-value
Apo-CIII0a	0.561	**0.009**	0.708	0.157
Apo-CIII0c	0.074	0.460	0.650	0.393
Apo-CIII1	0.303	0.630	**0.028**	**0.014**
Apo-CIII2	0.062	0.527	**0.038**	0.227

Bold for p<0.05
Within total DiaGene cohort

Statistical methods

The associations of apo-CIII glycosylation with lipid markers were analysed in the DiaGene Study using linear regression, separately for each apo-CIII proteoform. The proteoforms were considered the dependent variables. First, the analysis was performed with adjustments for age and sex; subsequently, a sensitivity analysis was performed for the use of lipid-lowering therapy (statins and fibrates as a binary variable). Next, this analysis was repeated for people with and without type 2 diabetes separately to investigate potential differences in medication effects. Since four separate proteoforms were analysed, the Bonferroni corrected p-value for significance was applied as $p<0.0125$ (0.05/4).

The genetic variants with a p-value $<1*10^{-6}$ in the GWASs were selected for further statistical analyses with lipid markers and prevalence of type 2 diabetes. We performed multiple linear regression analyses of the association of the genetic variants with lipid markers HDL-cholesterol (HDL-c), non-HDL-c, LDL-c, total cholesterol and triglycerides, adjusted for age and sex. The triglyceride concentrations were natural log-transformed for the statistical analyses; p-values are shown for log-transformed triglyceride levels, while effect sizes are shown for non-transformed triglyceride concentrations. Next, we performed logistic regression analyses for the associations of the genetic variants with the prevalence of type 2 diabetes, adjusted for age and sex. The Bonferroni corrected threshold for significance was calculated as $p<0.0045$ (0.05/11), as we analysed 11 single nucleotide polymorphisms (SNPs). Variants showing associations with a p-value $<1*10^{-6}$ in the GWASs were investigated for associations with lipid markers in the Hoorn DCS Study. The results of DiaGene and Hoorn DCS of these variants with clinical outcomes were meta-analysed using beta, standard error and number of participants of each cohort. Subsequently, SNPs were looked up in public GWAS databases Type 2 Diabetes Knowledge Portal[189] and IEU OpenGWAS[190] and in expression quantitative trait loci (eQTL) database eQTLGen[191] on May 25th 2022.

IBM SPSS 25.0 was used for all statistical analyses within the DiaGene study, R version 4.0.5 was used for statistical analysis in the Hoorn DCS and meta-analyses.

RESULTS

Cohort characteristics

Apo-CIII glycosylation was available for 2,318 participants in the DiaGene Study; the participants' mean age was 65.3 (SD 9.5) years, 51% was female, the median BMI was 28.0 (IQR 25.2 to 31.6) kg/m^2, 68% had type 2 diabetes, and 48% used statins or fibrates as lipid-lowering therapy. In the Hoorn DCS, genotypes of 5,409 participants with type 2 diabetes were available; the mean age was 61.1 (SD 11.0) years, 44.6% was female, the median BMI was 29.4 (IQR 26.6 to 33.1) kg/m^2 and 42.1% used statins or fibrates as lipid-lowering therapy (**Table 1**).

Table 1. General characteristics of the study populations.

	DiaGene	Hoorn DCS
Number of participants	2,318	5,409
Female sex, n (%)	1,175 (50.7)	2,414 (44.6)
Age, year, mean (±SD)	65.3 (9.5)	61.1 (11,0)
BMI, kg/m2, median (IQR)	28.0 (25.2-31.6)	29.4 (26.6-33.1)
Type 2 diabetes, n (%)	1,572 (67.8)	5,409 (100)
HDL-cholesterol, mmol/L, median (IQR)	1.2 (1.0-1.5)	1.2 (1-1.4)
Non-HDL cholesterol, mmol/L, mean (±SD)	3.4 (1.0)	3.8 (1.5)
LDL-cholesterol, mmol/L, mean (±SD)	2.8 (1.0)	2.9 (1.5)
Triglycerides, mmol/L, median (IQR)	1.4 (1.0-1.9)	1.7 (1.2-2.3)
Total cholesterol, mmol/L, mean (±SD)	4.7 (1.1)	5.0 (1.6)
Use of lipid lowering drugs, n (%)	1,113 (50.3)	2,276 (42.1)

Continuous data are presented as mean (and standard deviation) or median (and interquartile range) for normal and non-normal distributions respectively. The distribution of the clinical variables was considered normal when Skewness and Kurtosis were within the range of −1 to +1.

Apo-CIII *O*-glycosylation, type 2 diabetes and lipid parameters

Apo-CIII$_{0a}$ had a significant, negative association with the prevalence of type 2 diabetes (**Table 2**). In addition, apo-CIII$_{0a}$ was significantly associated with high cholesterol levels from both HDL and non-HDL particles, including LDL. Apo-CIII$_{0c}$ was associated with high triglycerides, total cholesterol, and non-HDL-c; but not with HDL-c or LDL-c. These associations point at more TRLs when the proportion of apo-CIII$_{0c}$ is high. Apo-CIII$_1$ and the apo-CIII$_1$/apo-CIII$_2$ ratio were positively and apo-CIII$_2$ was negatively associated with LDL-c, non-HDL-c, total cholesterol, and triglycerides. HDL-c was associated with a lower

apo-$CIII_1$/apo-$CIII_2$ ratio, lower proportion of apo-$CIII_1$, and higher proportion of apo-$CIII_2$ (**Table 2**). Adjustment for the use of lipid-lowering therapy and repeating these analyses for people with and without type 2 diabetes separately did not alter the direction of effect, although some associations lost significance (**Supplementary Table S2**).

Table 2. Associations of apo-CIII proteoforms with lipids and type 2 diabetes within the DiaGene study.

	HDL-c		
Proteoform	**beta**	**95% CI**	**p-value**
Apo-$CIII_{0a}$	0.007	0.005 to 0.009	$5.81*10^{-9}$
Apo-$CIII_{0c}$	-0.002	-0.005 to 0.001	0.232
Apo-$CIII_1$	-0.011	-0.015 to -0.007	$3.24*10^{-8}$
Apo-$CIII_2$	0.013	0.008 to 0.018	$3.20*10^{-7}$
Apo-$CIII_1$/apo-$CIII_2$ ratio	-0.321	-0.422 to -0.221	$4.00*10^{-10}$
	Non-HDL-c		
Proteoform	**beta**	**95% CI**	**p-value**
Apo-$CIII_{0a}$	0.004	0.003 to 0.004	$1.48*10^{-18}$
Apo-$CIII_{0c}$	0.002	0.001 to 0.003	$3.09*10^{-4}$
Apo-$CIII_1$	0.006	0.004 to 0.007	$9.04*10^{-16}$
Apo-$CIII_2$	-0.007	-0.009 to -0.006	$3.60*10^{-17}$
Apo-$CIII_1$/apo-$CIII_2$ ratio	0.133	0.099 to 0.167	$2.82*10^{-14}$
	LDL-c		
Proteoform	**beta**	**95% CI**	**p-value**
Apo-$CIII_{0a}$	0.004	0.003 to 0.005	$3.86*10^{-20}$
Apo-$CIII_{0c}$	0.001	0.000 to 0.002	0.041
Apo-$CIII_1$	0.004	0.003 to 0.005	$4.33*10^{-8}$
Apo-$CIII_2$	-0.005	-0.007 to -0.003	$5.74*10^{-8}$
Apo-$CIII_1$/apo-$CIII_2$ ratio	0.080	0.045 to 0.116	$9.65*10^{-6}$
	Total cholesterol		
Proteoform	**beta**	**95% CI**	**p-value**
Apo-$CIII_{0a}$	0.004	0.003 to 0.005	$4.37*10^{-24}$
Apo-$CIII_{0c}$	0.002	0.001 to 0.003	0.003
Apo-$CIII_1$	0.004	0.002 to 0.005	$8.99*10^{-9}$
Apo-$CIII_2$	-0.005	-0.007 to -0.004	$4.17*10^{-10}$
Apo-$CIII_1$/apo-$CIII_2$ ratio	0.084	0.052 to 0.116	$2.99*10^{-7}$
	Triglycerides		
Proteoform	**beta**	**95% CI**	**p-value**
Apo-$CIII_{0a}$	0.001	0.000 to 0.001	0.045
Apo-$CIII_{0c}$	0.003	0.002 to 0.004	$2.08*10^{-8}$
Apo-$CIII_1$	0.008	0.007 to 0.009	$9.66*10^{-49}$
Apo-$CIII_2$	-0.011	-0.012 to -0.009	$1.03*10^{-49}$
Apo-$CIII_1$/apo-$CIII_2$ ratio	0.225	0.193 to 0.257	$2.26*10^{-51}$
	Type 2 diabetes		
Proteoform	**Beta**	**95% CI**	**p-value**
Apo-$CIII_{0a}$	-29.59	-34.43 to -24.74	$5.42*10^{-33}$
Apo-$CIII_{0c}$	-0.963	-4.338 to 2.412	0.576
Apo-$CIII_1$	1.275	-1.399 to 3.950	0.350
Apo-$CIII_2$	-0.296	-2.377 to 1.786	0.781
Apo-$CIII_1$/apo-$CIII_2$ ratio	0.110	0.002 to 0.217	0.046

Adjusted for age and sex. Apo-$CIII_{0a}$ normalised to all four proteoforms: apo-$CIII_{0a}$, apo-$CIII_{0c}$, apo-$CIII_1$ and apo-$CIII_2$. The sum of the glycoforms Apo-$CIII_{0c}$, Apo-$CIII_1$, and Apo-$CIII_2$ was set to 1.0.

Table S2. Associations of apolipoprotein-CIII proteoforms with lipids in people with and without T2D

DiaGene

HDL-c	Adjusted for age and sex					Adjusted for age, sex and lipid lowering medication														
	Total DiaGene cohort					Total DiaGene cohort					T2D					Without T2D				
	n=	2182				n=	2081				n=	1399				n=	682			
			95% C.I.for Beta					95% C.I.for Beta					95% C.I.for Beta					95% C.I.for Beta		
trait	beta	SE	Lower	Upper	p-value	beta	SE	Lower	Upper	p-value	beta	SE	Lower	Upper	p-value	beta	SE	Lower	Upper	p-value
Apo-CIII0a	0.007	0.001	0.005	0.009	**5.81E-09**	0.004	0.001	0.002	0.007	**0.001**	0.002	0.001	-0.001	0.005	0.248	0.001	0.002	-0.003	0.006	0.577
Apo-CIII0c	-0.002	0.002	-0.005	0.001	0.232	-0.002	0.002	-0.005	0.002	0.292	-0.002	0.002	-0.007	0.002	0.343	-0.002	0.003	-0.008	0.004	0.527
Apo-CIII1	-0.011	0.002	-0.015	-0.007	**3.24E-08**	-0.013	0.002	-0.017	-0.009	**2.52E-09**	-0.012	0.003	-0.018	-0.006	**2.78E-05**	-0.012	0.003	-0.019	-0.005	**3.77E-04**
Apo-CIII2	0.013	0.003	0.008	0.018	**3.20E-07**	0.015	0.003	0.009	0.020	**9.31E-08**	0.015	0.004	0.007	0.022	**1.06E-04**	0.014	0.004	0.006	0.023	**1.10E-03**
Apo-CIII1/apo-CIII2 ratio	-0.321	0.051	-0.422	-0.221	**4.00E-10**	-0.348	0.054	-0.453	-0.243	**1.00E-10**	-0.321	0.073	-0.465	-0.177	**1.30E-05**	-0.314	0.084	-0.478	-0.151	**1.83E-04**

non-HDL-c	Adjusted for age and sex					Adjusted for age, sex and lipid lowering medication														
	Total DiaGene cohort					Total DiaGene cohort					T2D					Without T2D				
	n=	2182				n=	2081				n=	1399				n=	682			
			95% C.I.for Beta					95% C.I.for Beta					95% C.I.for Beta					95% C.I.for Beta		
trait	beta	SE	Lower	Upper	p-value	beta	SE	Lower	Upper	p-value	beta	SE	Lower	Upper	p-value	beta	SE	Lower	Upper	p-value
Apo-CIII0a	0.004	0.000	0.003	0.004	**1.48E-18**	0.002	0.000	0.002	0.003	**1.50E-07**	0.002	0.001	0.001	0.003	**9.51E-04**	0.001	0.001	0.000	0.003	0.145
Apo-CIII0c	0.002	0.001	0.001	0.003	**3.09E-04**	0.003	0.001	0.002	0.004	**1.18E-05**	0.003	0.001	0.002	0.005	**3.53E-05**	0.002	0.001	0.000	0.004	0.063
Apo-CIII1	0.006	0.001	0.004	0.007	**9.04E-16**	0.007	0.001	0.005	0.008	**7.05E-17**	0.008	0.001	0.005	0.010	**8.21E-13**	0.007	0.001	0.005	0.010	**4.89E-09**
Apo-CIII2	-0.007	0.001	-0.009	-0.006	**3.60E-17**	-0.009	0.001	-0.011	-0.007	**2.21E-19**	-0.011	0.001	-0.013	-0.008	**8.47E-15**	-0.009	0.002	-0.012	-0.006	**4.78E-09**
Apo-CIII1/apo-CIII2 ratio	0.133	0.017	0.099	0.167	**2.82E-14**	0.165	0.020	0.126	0.204	**1.32E-16**	0.199	0.027	0.146	0.251	**1.44E-13**	0.187	0.030	0.128	0.247	**1.27E-09**

DiaGene

LDL-c	Adjusted for age and sex					Adjusted for age, sex and lipid lowering medication														
	Total DiaGene cohort					Total DiaGene cohort					T2D					Without T2D				
	n=	2161				n=	2061				n=	1378				n=	683			
			95% C.I.for Beta					95% C.I.for Beta					95% C.I.for Beta					95% C.I.for Beta		
trait	beta	SE	Lower	Upper	p-value	beta	SE	Lower	Upper	p-value	beta	SE	Lower	Upper	p-value	beta	SE	Lower	Upper	p-value
Apo-CIII0a	0.004	0.000	0.003	0.005	**3.86E-20**	0.003	0.000	0.002	0.004	**2.22E-07**	0.002	0.001	0.000	0.003	**0.015**	0.002	0.001	0.000	0.003	0.118
Apo-CIII0c	0.001	0.001	0.000	0.002	**0.041**	0.002	0.001	0.001	0.003	**0.004**	0.002	0.001	0.000	0.004	**0.038**	0.002	0.001	0.000	0.005	**0.039**
Apo-CIII1	0.004	0.001	0.003	0.005	**4.33E-08**	0.005	0.001	0.003	0.006	**2.47E-08**	0.005	0.001	0.003	0.008	**1.03E-05**	0.007	0.001	0.005	0.010	**1.13E-07**
Apo-CIII2	-0.005	0.001	-0.007	-0.003	**5.74E-08**	-0.006	0.001	-0.009	-0.004	**4.81E-09**	-0.007	0.002	-0.010	-0.004	**1.52E-05**	-0.010	0.002	-0.013	-0.006	**2.68E-08**
Apo-CIII1/apo-CIII2 ratio	0.080	0.018	0.045	0.116	**9.65E-06**	0.104	0.022	0.062	0.147	**1.29E-06**	0.115	0.030	0.057	0.174	**1.25E-04**	0.189	0.033	0.124	0.254	**1.70E-08**

DiaGene

Total cholesterol	Adjusted for age and sex					Adjusted for age, sex and lipid lowering medication														
	Total DiaGene cohort					Total DiaGene cohort					T2D					Without T2D				
	n=	2185				n=	2084				n=	1402				n=	682			
			95% C.I.for Beta					95% C.I.for Beta					95% C.I.for Beta					95% C.I.for Beta		
trait	beta	SE	Lower	Upper	p-value	beta	SE	Lower	Upper	p-value	beta	SE	Lower	Upper	p-value	beta	SE	Lower	Upper	p-value
Apo-CIII0a	0.004	0.000	0.003	0.005	**4.37E-24**	0.003	0.000	0.002	0.004	**2.80E-10**	0.002	0.001	0.001	0.003	**2.50E-04**	0.002	0.001	0.000	0.003	0.090
Apo-CIII0c	0.002	0.001	0.001	0.003	**0.003**	0.002	0.001	0.001	0.004	**9.70E-05**	0.003	0.001	0.001	0.005	**1.58E-04**	0.002	0.001	0.000	0.004	0.098
Apo-CIII1	0.004	0.001	0.002	0.005	**8.99E-09**	0.005	0.001	0.003	0.006	**3.13E-09**	0.006	0.001	0.004	0.008	**2.79E-08**	0.006	0.001	0.003	0.008	**4.59E-06**
Apo-CIII2	-0.005	0.001	-0.007	-0.004	**4.17E-10**	-0.007	0.001	-0.009	-0.005	**1.11E-11**	-0.008	0.001	-0.011	-0.006	**3.57E-10**	-0.008	0.002	-0.011	-0.004	**2.63E-06**
Apo-CIII1/apo-CIII2 ratio	0.084	0.016	0.052	0.116	**2.99E-07**	0.110	0.019	0.072	0.148	**1.31E-08**	0.152	0.026	0.101	0.204	**9.94E-09**	0.149	0.031	0.088	0.211	**2.22E-06**

DiaGene

Triglycerides	Adjusted for age and sex					Adjusted for age, sex and lipid lowering medication														
	Total DiaGene cohort					Total DiaGene cohort					T2D					Without T2D				
	n=	2184				n=	2083				n=	1400				n=	683			
			95% C.I.for Beta					95% C.I.for Beta					95% C.I.for Beta					95% C.I.for Beta		
trait	beta	SE	Lower	Upper	p-value	beta	SE	Lower	Upper	p-value	beta	SE	Lower	Upper	p-value	beta	SE	Lower	Upper	p-value
Apo-CIII0a	0.001	0.000	0.000	0.001	**0.045**	0.001	0.000	0.000	0.002	**3.67E-04**	0.001	0.000	0.001	0.002	**1.39E-04**	0.003	0.001	0.000	0.005	**0.008**
Apo-CIII0c	0.003	0.001	0.002	0.004	**2.08E-08**	0.003	0.001	0.002	0.004	**6.78E-08**	0.003	0.001	0.002	0.004	**9.25E-08**	0.002	0.002	-0.001	0.005	0.096
Apo-CIII1	0.008	0.001	0.007	0.009	**9.66E-49**	0.008	0.001	0.007	0.009	**3.75E-48**	0.007	0.001	0.005	0.008	**3.94E-29**	0.015	0.002	0.012	0.019	**2.10E-21**
Apo-CIII2	-0.011	0.001	-0.012	-0.009	**1.03E-49**	-0.011	0.001	-0.012	-0.009	**9.45E-48**	-0.010	0.001	-0.012	-0.008	**8.27E-32**	-0.017	0.002	-0.022	-0.013	**2.08E-17**
Apo-CIII1/apo-CIII2 ratio	0.225	0.016	0.193	0.257	**2.26E-51**	0.228	0.017	0.194	0.261	**1.64E-49**	0.205	0.019	0.168	0.242	**3.57E-31**	0.359	0.042	0.276	0.442	**4.79E-19**

Bold for p<0.05; **bold and underlined** for p<0.05/4=0.0125

GWASs of apo-CIII *O*-glycosylation proteoforms

We conducted GWASs on the four main proteoforms of apo-CIII and the apo-$CIII_1$/apo-$CIII_2$ ratio in the DiaGene study. In total, we identified 11 SNPs with $p<1*10^{-6}$. The GWAS results are summarised in **Table 3, Figure 1 and Supplementary Figure S2**. One signal, the A allele of rs4846913 at the *GALNT2* locus on chromosome 1, was genome-wide significantly associated with decreased apo-$CIII_{0a}$ ($p=9.77*10^{-36}$) (**Table 3, Supplementary Figure S3**). This SNP was also associated with high HDL-c (p=0.065) and low triglycerides (p=0.006) in the meta-analysis (**Table 4**), in line with public GWAS data ($p= 7.46*10^{-146\ 190}$ and $2.28*10^{-235\ 189}$ respectively). The association of rs4846913 with apo-$CIII_{0a}$ remained genome-wide significant when analysed for people with and without type 2 diabetes separately (**Supplementary Table S5**). The variant rs35498929-T, also in the *GALNT2* locus, was associated at a suggestive significance level with decreased apo-$CIII_{0a}$ ($p=7.44*10^{-7}$) and with higher prevalence of type 2 diabetes (p=0.006) in the relatively small DiaGene study for such analyses (**Supplementary Figure S3 and Supplementary Table S3**).

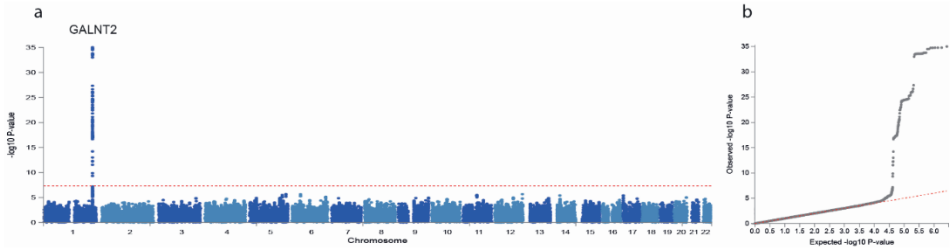


Figure 1. Manhattan and QQ plot of apo-CIII$_{0a}$ GWAS. GWAS of apo-CIII$_{0a}$ showing associations with alleles at the *GALNT2* locus on chromosome 1. Manhattan plot (left) showing significance of the association of each SNP allele with apo-CIII$_{0a}$ by plotting the -log10 of the p-value against the genomic position. The horizontal red dotted line corresponds to the genome-wide significance threshold of $p=5*10^{-8}$. Quantile-quantile plot (right) is a plot of the observed $-\log_{10}(p)$ against the expected $-\log_{10}(p)$ under the null hypothesis of no association. Deviation above the red dotted y=x line indicates lower p-values that would be expected to occur by chance and implies statistically significant association.

Table 3. Loci associated with at least one apo-CIII proteoform in the DiaGene study.

Proteoform	rsID	Chr	Pos	EA	RA	EAF	Beta	SE	*p*-Value	Locus
Apo-$CIII_{0a}$	rs35498929	1	230286016	T	C	0.1249	−0.1223	0.02472	7.44×10^{-7}	*GALNT2* (intronic)
	rs4846913	1	230294715	A	C	0.5964	−0.2027	0.01624	9.77×10^{-36}	*GALNT2* (intronic)
	rs3213497	1	230416320	T	C	0.1282	−0.1271	0.02407	1.30×10^{-7}	*GALNT2:RP5-956O18.3* (ncRNA exonic)
Apo-$CIII_{0c}$	rs9378785	6	3316862	C	T	0.0508	−0.0918	0.01705	7.34×10^{-8}	*SLC22A23* (intronic)
Apo-$CIII_1$	rs10842926	12	27689893	C	G	0.0826	−0.0145	0.00292 8	7.07×10^{-7}	*PPFIBP1* (intronic)
Apo-$CIII_2$	rs67086575	2	27686480	G	A	0.1502	−0.0469	0.00947 3	7.32×10^{-7}	*IFT172* (intronic)
	rs2493926	6	148614267	C	T	0.1112	0.0520	0.01058	9.12×10^{-7}	*SASH1* (intronic)
	rs2481968	13	28567172	C	A	0.4874	0.0334	0.00664 7	4.90×10^{-7}	*RN7SL272P* (intergenic)
	rs7175584	15	97500494	T	C	0.5638	−0.0339	0.00671	4.48×10^{-7}	*RN7SKP181* (intergenic)
	rs10412211	19	13591542	T	G	0.4579	−0.0336	0.00671 3	5.62×10^{-7}	*CACNA1A* (intronic)
Apo-$CIII_1$/apo-$CIII_2$ ratio	rs67086575	2	27686480	G	A	0.1502	0.0561	0.01133	7.37×10^{-7}	*IFT172* (intronic)
	rs9462715	6	12368221	C	A	0.0571	−0.0826	0.01676	8.35×10^{-7}	*RN7SKP293* (intergenic)
	rs2493926	6	148614267	C	T	0.1112	−0.0627	0.01266	7.46×10^{-7}	*SASH1* (intronic)
	rs7175584	15	97500494	T	C	0.5638	0.0419	0.00802 8	1.80×10^{-7}	*RN7SKP181* (intergenic)
	rs10412211	19	13591542	T	G	0.4579	0.0396	0.00803 1	8.24×10^{-7}	*CACNA1A* (intronic)

An association was considered significant if the p value was lower than or equal to $5*10^{-8}$, the genome-wide significance threshold. Associations with p-values lower than or equal to $1*10^{-6}$ were considered at suggestive significance level. Chr, chromosome; Pos, position; EA, effect allele; RA, reference allele; EAF, effect allele frequency; SE, standard error; ncRNA, non-coding RNA.

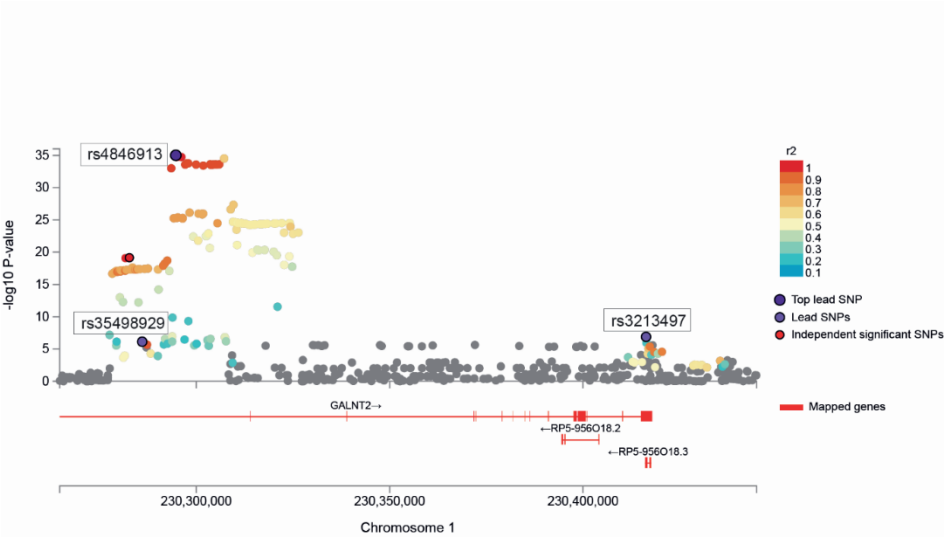


Figure S3. Region plot for the A allele of rs4846913 at the *GALNT2* locus on chromosome 1.

Table S3. Associations of leadSNPs with T2D.

						95% C.I.for Beta		
uniqID	rsID (dbSNP build 146)	effect allele	reference allele	Beta	SE	Lower	Upper	p-value
1:230286016:C:T	rs35498929	T	C	0,341	0,123	0,099	0,582	**0,006**
1:230294715:A:C	rs4846913	A	C	0,033	0,078	-0,120	0,185	0,676
1:230416320:C:T	rs3213497	T	C	-0,020	0,115	-0,245	0,206	0,864
2:27686480:A:G	rs67086575	G	A	-0,143	0,105	-0,348	0,063	0,173
6:3316862:C:T	rs9378785	C	T	0,283	0,197	-0,103	0,670	0,151
6:12368221:A:C	rs9462715	C	A	-0,180	0,168	-0,510	0,150	0,285
6:148614267:C:T	rs2493926	C	T	0,163	0,128	-0,088	0,413	0,203
12:27689893:C:G	rs10842926	C	G	-0,145	0,138	-0,414	0,125	0,293
13:28567172:A:C	rs2481968	C	A	-0,072	0,076	-0,221	0,077	0,344
15:97500494:C:T	rs7175584	T	C	0,177	0,077	0,025	0,328	**0,022**
19:13591542:G:T	rs10412211	T	G	0,071	0,078	-0,082	0,224	0,364

Bold for p<0.05
Within the DiaGene cohort

n = 1759

Two variants reached a suggestive significance level for associations with one or more apo-CIII proteoforms without reaching formal genome-wide significance but were also associated with one or more lipid traits in the meta-analyses. The rs10842926-C in the *PPFIBP1* locus, was associated with decreased apo-$CIII_1$ ($p=7.07*10^{-07}$, **Supplementary figure S4, Table 3**). This variant was also associated with high LDL-c (p=0.016) and low triglycerides (p=0.039) with nominal significance in meta-analysis (**Table 4**) but was not significant in the Type 2 Diabetes Knowledge Portal[189]. The rs67086575-G allele, in the intronic region of the *IFT172*-gene was associated with apo-$CIII_2$ and the apo-$CIII_1$/apo-$CIII_2$ ratio ($p=7.32*10^{-07}$ and $p=7.37*10^{-07}$, respectively); and was significant for higher triglycerides ($p=1.07*10^{-5}$) in the meta-analysis, and for higher LDL-c, total cholesterol, and non-HDL-c (p=0.019, p=0.025, and p=0.013 respectively) in the DiaGene study (**Supplementary figure S5, Table 3**). The associations with triglycerides ($p=4.84*10^{-287}$)[189], LDL-c ($p=2.90*10^{-51}$)[190], and total cholesterol ($p= 4.76*10^{-26}$)[190]were in line with public data from the Type 2 Diabetes Knowledge Portal and the IEU OpenGWAS Project. The associations of our 11 identified genetic loci with apo-CIII proteoforms had the same direction of effect in people with and without type 2 diabetes (**Supplementary Table S5**). Other variants had a promising association with an apo-CIII proteoform and showed a significant association with clinical traits in one of the studies, but these signals did not survive the meta-analyses (in **Supplementary Tables S3 and S4**).

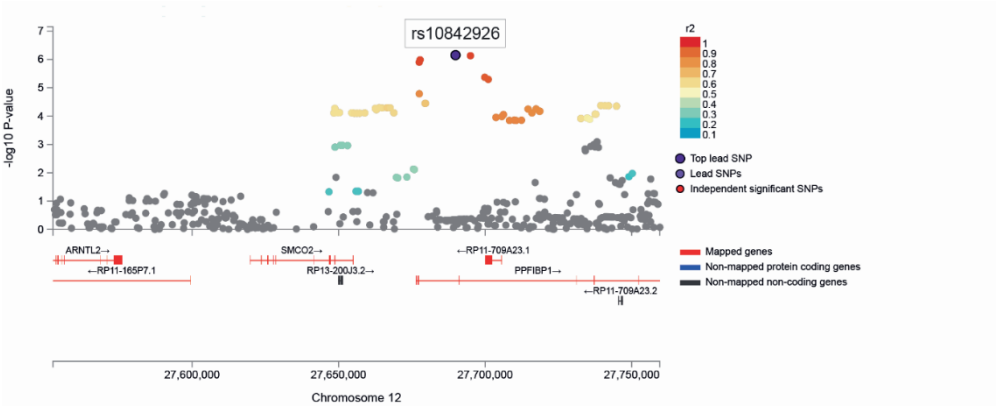


Figure S4. Region plot for the rs10842926-C in the *PPFIBP1* locus.

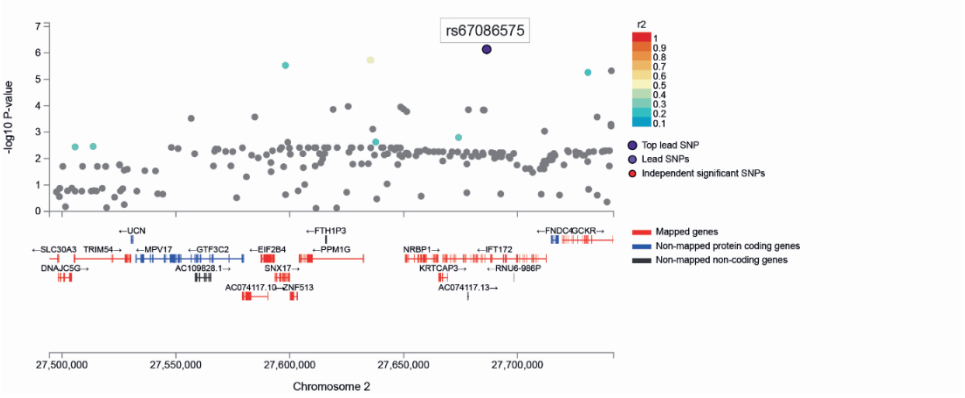


Figure S5. Region plot for the rs67086575-G allele in the intronic region of the IFT172-gene.

Table 4. Significant associations of genetic variants with lipids.

HDL-c		***DiaGene***			***Hoorn DCS***			***Meta-analysis***
rsID	EA	beta	95% C.I.	p-value	beta	95% C.I.	p-value	p-value
rs4846913	A	0.006	-0.017 to 0.030	0.606	0.011	-0.001 to 0.023	0.076	0.065
rs9378785	C	-0.058	-0.113 to -0.003	0.040	0.005	-0.025 to 0.033	0.751	0.472
Non-HDL-c		***DiaGene***			***Hoorn DCS***			***Meta-analysis***
rsID	EA	beta	95% C.I.	p-value	beta	95% C.I.	p-value	p-value
rs67086575	G	0.120	0.025 to 0.215	0.013	0.001	-0.056 to 0.059	0.961	0.207
rs2493926	C	-0.157	-0.266 to -0.048	0.005	0.030	-0.041 to 0.101	0.412	0.522
LDL-c		***DiaGene***			***Hoorn DCS***			***Meta-analysis***
rsID	EA	beta	95% C.I.	p-value	beta	95% C.I.	p-value	p-value
rs67086575	G	0.109	0.018 to 0.201	0.019	-0.045	-0.097 to 0.006	0.086	0.707
rs2493926	C	-0.183	-0.288 to -0.078	0.001	0.035	-0.029 to 0.099	0.289	0.464
rs10842926	C	0.038	-0.084 to 0.160	0.547	0.089	0.016 to 0.162	0.016	0.016
Total cholesterol		***DiaGene***			***Hoorn DCS***			***Meta-analysis***
rsID	EA	beta	95% C.I.	p-value	beta	95% C.I.	p-value	p-value
rs67086575	G	0.115	0.014 to 0.216	0.025	0.007	-0.051 to 0.065	0.817	0.190
rs2493926	C	-0.171	-0.286 to -0.055	0.004	0.017	-0.055 to 0.088	0.650	0.311
Triglycerides		***DiaGene***			***Hoorn DCS***			***Meta-analysis***
rsID	EA	beta	95% C.I.	p-value	beta	95% C.I.	p-value	p-value
rs4846913	A	-0.063	-0.135 to 0.008	0.032	-0.052	-0.097 to -0.006	0.034	0.006
rs67086575	G	0.084	-0.014 to 0.183	0.017	0.084	0.023 to 0.145	$2.20*10^{-4}$	$1.07*10^{-5}$
rs10842926	C	-0.081	-0.212 to 0.051	0.061	-0.061	-0.147 to 0.025	0.185	0.039
rs2481968	C	-0.076	-0.146 to -0.007	0.025	0.017	-0.028 to 0.062	0.759	0.391

Adjusted for age and sex. EA, effect allele.

Table S4. Associations of leadSNPs with lipids.

HDL-c				***DiaGene***					***Hoorn DCS***					***Meta-analysis***
				n= 1656					n= 5304					
						95% C.I.for Beta					95% C.I.for Beta			
uniqID	rsID (dbSNP build 146)	EA	RE	beta	SE	Lower	Upper	p-value	beta	SE	Lower	Upper	p-value	p-value
1:230286016:C:T	rs35498929	T	C	-0.032	0.018	-0.067	0.003	0.076	0.005	0.009	-0.013	0.023	0.580	0.702
1:230294715:A:C	rs4846913	A	C	0.006	0.012	-0.017	0.030	0.606	0.011	0.006	-0.001	0.023	0.076	0.065
1:230416320:C:T	rs3213497	T	C	0.013	0.018	-0.021	0.048	0.448	0.003	0.009	-0.014	0.020	0.733	0.520
2:27686480:A:G	rs67086575	G	A	-0.004	0.017	-0.037	0.028	0.801	0.006	0.008	-0.011	0.022	0.497	0.589
6:3316862:C:T	rs9378785	C	T	-0.058	0.028	-0.113	-0.003	**0.040**	0.005	0.015	-0.024	0.033	0.751	0.472
6:12368221:A:C	rs9462715	C	A	0.008	0.027	-0.045	0.060	0.767	0.007	0.014	-0.021	0.036	0.603	0.561
6:148614267:C:T	rs2493926	C	T	-0.013	0.019	-0.050	0.024	0.487	-0.011	0.010	-0.031	0.009	0.274	0.196
12:27689893:C:G	rs10842926	C	G	0.025	0.022	-0.018	0.068	0.259	0.009	0.012	-0.014	0.031	0.451	0.227
13:28567172:A:C	rs2481968	C	A	0.001	0.012	-0.022	0.024	0.937	0.008	0.006	-0.004	0.020	0.171	0.228
15:97500494:C:T	rs7175584	T	C	-0.021	0.012	-0.044	0.003	0.084	0.011	0.006	0.000	0.023	0.060	0.227
19:13591542:G:T	rs10412211	T	G	-0.009	0.012	-0.033	0.014	0.449	0.004	0.006	-0.008	0.016	0.528	0.829
non-HDL-c				***DiaGene***					***Hoorn DCS***					***Meta-analysis***
				n= 1656					n= 5304					
						95% C.I.for Beta					95% C.I.for Beta			
uniqID	rsID (dbSNP build 146)	EA	RE	beta	SE	Lower	Upper	p-value	beta	SE	Lower	Upper	p-value	p-value
1:230286016:C:T	rs35498929	T	C	-0.058	0.053	-0.161	0.045	0.267	0.051	0.033	-0.013	0.115	0.121	0.413
1:230294715:A:C	rs4846913	A	C	-0.042	0.035	-0.112	0.027	0.231	-0.013	0.022	-0.056	0.030	0.545	0.265

1:230416320:C:T	rs3213497	T	C	0.002	0.052	-0.100	0.105	0.963	-0.058	0.032	-0.120	0.004	0.066	0.118
2:27686480:A:G	rs67086575	G	A	0.120	0.048	0.025	0.215	**0.013**	0.001	0.029	-0.056	0.059	0.961	0.207
6:3316862:C:T	rs9378785	C	T	-0.110	0.083	-0.272	0.053	0.187	-0.052	0.052	-0.154	0.050	0.319	0.129
6:12368221:A:C	rs9462715	C	A	-0.032	0.079	-0.186	0.122	0.685	-0.004	0.051	-0.104	0.096	0.937	0.790
6:148614267:C:T	rs2493926	C	T	-0.157	0.056	-0.266	-0.048	**0.005**	0.030	0.036	-0.041	0.101	0.412	0.522
12:27689893:C:G	rs10842926	C	G	0.011	0.065	-0.116	0.138	0.865	0.073	0.041	-0.008	0.155	0.077	0.102
13:28567172:A:C	rs2481968	C	A	-0.052	0.034	-0.119	0.015	0.130	0.022	0.022	-0.020	0.065	0.304	0.899
15:97500494:C:T	rs7175584	T	C	-0.009	0.035	-0.078	0.059	0.787	-0.002	0.022	-0.045	0.040	0.910	0.838
19:13591542:G:T	rs10412211	T	G	0.023	0.035	-0.046	0.092	0.508	0.024	0.022	-0.019	0.067	0.269	0.203
LDL-c						***DiaGene***					***Hoorn DCS***			***Meta-analysis***
				n=	1638				n=	5213				
						95% C.I.for Beta					95% C.I.for Beta			
uniqID	rsID (dbSNP build 146)	EA	RE	beta	SE	Lower	Upper	p-value	beta	SE	Lower	Upper	p-value	p-value
1:230286016:C:T	rs35498929	T	C	-0.055	0.050	-0.154	0.044	0.273	0.053	0.029	-0.005	0.110	0.075	0.291
1:230294715:A:C	rs4846913	A	C	-0.023	0.034	-0.090	0.044	0.505	0.011	0.020	-0.027	0.050	0.570	0.882
1:230416320:C:T	rs3213497	T	C	0.002	0.050	-0.096	0.100	0.967	-0.034	0.028	-0.089	0.022	0.234	0.298
2:27686480:A:G	rs67086575	G	A	0.109	0.047	0.018	0.201	**0.019**	-0.045	0.026	-0.097	0.006	0.086	0.707
6:3316862:C:T	rs9378785	C	T	-0.019	0.080	-0.176	0.137	0.810	-0.053	0.047	-0.146	0.039	0.256	0.271
6:12368221:A:C	rs9462715	C	A	0.008	0.076	-0.141	0.156	0.921	0.025	0.046	-0.065	0.115	0.584	0.599
6:148614267:C:T	rs2493926	C	T	-0.183	0.054	-0.288	-0.078	**0.001**	0.035	0.033	-0.029	0.099	0.289	0.464
12:27689893:C:G	rs10842926	C	G	0.038	0.062	-0.084	0.160	0.547	0.089	0.037	0.016	0.162	**0.016**	**0.016**
13:28567172:A:C	rs2481968	C	A	-0.020	0.033	-0.085	0.044	0.537	0.020	0.020	-0.018	0.059	0.299	0.565
15:97500494:C:T	rs7175584	T	C	-0.028	0.034	-0.094	0.038	0.408	0.016	0.019	-0.022	0.055	0.397	0.740
19:13591542:G:T	rs10412211	T	G	0.034	0.034	-0.032	0.100	0.316	0.015	0.020	-0.023	0.053	0.445	0.253

Total cholesterol				***DiaGene***					***Hoorn DCS***					***Meta-analysis***
				n= 1658					n= 5309					
						95% C.I.for Beta					95% C.I.for Beta			
uniqID	rsID (dbSNP build 146)	EA	RE	beta	SE	Lower	Upper	p-value	beta	SE	Lower	Upper	p-value	p-value
1:230286016:C:T	rs35498929	T	C	-0.090	0.056	-0.200	0.019	0.105	0.058	0.033	-0.006	0.123	0.076	0.453
1:230294715:A:C	rs4846913	A	C	-0.036	0.038	-0.110	0.038	0.339	-0.003	0.022	-0.046	0.040	0.891	0.561
1:230416320:C:T	rs3213497	T	C	0.015	0.055	-0.093	0.124	0.780	-0.056	0.032	-0.118	0.006	0.079	0.163
2:27686480:A:G	rs67086575	G	A	0.115	0.051	0.014	0.216	**0.025**	0.007	0.029	-0.051	0.065	0.817	0.190
6:3316862:C:T	rs9378785	C	T	-0.168	0.088	-0.341	0.005	0.057	-0.050	0.052	-0.152	0.052	0.338	0.077
6:12368221:A:C	rs9462715	C	A	-0.024	0.083	-0.188	0.139	0.770	0.001	0.051	-0.099	0.102	0.979	0.901
6:148614267:C:T	rs2493926	C	T	-0.171	0.059	-0.286	-0.055	**0.004**	0.017	0.037	-0.055	0.088	0.650	0.311
12:27689893:C:G	rs10842926	C	G	0.036	0.069	-0.099	0.171	0.605	0.082	0.042	0.001	0.164	**0.047**	0.050
13:28567172:A:C	rs2481968	C	A	-0.052	0.036	-0.123	0.020	0.156	0.031	0.022	-0.012	0.074	0.164	0.599
15:97500494:C:T	rs7175584	T	C	-0.030	0.037	-0.103	0.043	0.424	0.009	0.022	-0.033	0.052	0.667	0.969
19:13591542:G:T	rs10412211	T	G	0.014	0.037	-0.059	0.087	0.709	0.028	0.022	-0.015	0.071	0.204	0.195
Triglycerides				***DiaGene***					***Hoorn DCS***					***Meta-analysis***
				n= 1658					n= 5307					
						95% C.I.for Beta					95% C.I.for Beta			
uniqID	rsID (dbSNP build 146)	EA	RE	beta	SE	Lower	Upper	p-value	beta	SE	Lower	Upper	p-value	p-value
1:230286016:C:T	rs35498929	T	C	0.018	0.054	-0.089	0.124	0.371	-0.003	0.035	-0.071	0.065	0.800	0.484
1:230294715:A:C	rs4846913	A	C	-0.063	0.037	-0.135	0.008	**0.032**	-0.052	0.023	-0.097	-0.006	**0.034**	**0.006**
1:230416320:C:T	rs3213497	T	C	-0.034	0.054	-0.140	0.071	0.719	-0.060	0.033	-0.126	0.005	0.100	0.108
2:27686480:A:G	rs67086575	G	A	0.084	0.050	-0.014	0.183	**0.017**	0.084	0.031	0.023	0.145	**2.20E-04**	**1.07E-05**

6:3316862:C:T	rs9378785	C	T	-0.098	0.086	-0.266	0.071	0.378	-0.002	0.055	-0.110	0.106	0.929	0.613
6:12368221:A:C	rs9462715	C	A	-0.068	0.081	-0.228	0.091	0.396	-0.056	0.054	-0.162	0.051	0.309	0.199
6:148614267:C:T	rs2493926	C	T	0.004	0.058	-0.109	0.117	0.809	-0.006	0.039	-0.081	0.070	0.620	0.753
12:27689893:C:G	rs10842926	C	G	-0.081	0.067	-0.212	0.051	0.061	-0.061	0.044	-0.147	0.025	0.185	**0.039**
13:28567172:A:C	rs2481968	C	A	-0.076	0.035	-0.146	-0.007	**0.025**	0.017	0.023	-0.028	0.062	0.759	0.391
15:97500494:C:T	rs7175584	T	C	0.049	0.036	-0.022	0.120	0.071	-0.037	0.023	-0.081	0.008	0.282	0.996
19:13591542:G:T	rs10412211	T	G	0.035	0.036	-0.037	0.106	0.432	0.005	0.023	-0.040	0.051	0.954	0.641

Bold for p<0.05, **bold and underlined** for p<0.05/11=0.0045

Table S5. Loci associated with at least one apo-CIII proteoform in the DiaGene study, separate for cases with and without type 2 diabetes

			Type 2 diabetes			No diabetes			
Proteoform	rsID	EA	Beta	SE	P-value	Beta	SE	P-value	Locus
Apo-CIII0a	rs35498929	T	-0.131	0.046	0.005	-0.104	0.028	1.85*10-4	GALNT2 (intronic)
	rs4846913	A	-0.219	0.027	3.05*10-15	-0.200	0.019	1.49*10-25	GALNT2 (intronic)
	rs3213497	T	-0.167	0.041	5.63*10-05	-0.100	0.029	0.001	GALNT2:RP5-956O18.3 (ncRNA exonic)
Apo-CIII0c	rs9378785	C	-0.069	0.037	0.060	-0.089	0.020	1.30*10-5	SLC22A23 (intronic)
Apo-CIII1	rs10842926	C	-0.016	0.005	0.001	-0.013	0.004	0.001	PPFIBP1 (intronic)
Apo-CIII2	rs67086575	G	-0.063	0.016	6.41*10-5	-0.039	0.011	8.01*10-4	IFT172 (intronic)
	rs2493926	C	0.026	0.020	0.198	0.051	0.013	7.33*10-5	SASH1 (intronic)
	rs2481968	C	0.027	0.011	0.018	0.033	0.008	5.09*10-5	RN7SL272P (intergenic)
	rs7175584	T	-0.017	0.012	0.138	-0.039	0.008	2.89*10-6	RN7SKP181 (intergenic)
	rs10412211	T	-0.029	0.012	0.013	-0.028	0.008	0.001	CACNA1A (intronic)
Apo-CIII1 / apo-CIII2 ratio	rs67086575	G	0.072	0.019	1.35*10-4	0.047	0.014	6.11*10-4	IFT172 (intronic)
	rs9462715	C	-0.042	0.029	0.156	-0.073	0.023	0.001	RN7SKP293 (intergenic)
	rs2493926	C	-0.030	0.024	0.204	-0.061	0.016	7.88*10-5	SASH1 (intronic)
	rs7175584	T	0.024	0.014	0.083	0.047	0.010	2.02*10-6	RN7SKP181 (intergenic)
	rs10412211	T	0.036	0.014	0.010	0.032	0.010	0.002	CACNA1A (intronic)

An association was considered significant if the p value was lower than or equal to 5*10−8, the genome-wide significance threshold. Associations with p-values lower than or equal to 1*10-6 were considered at suggestive significance level. Chr, chromosome; Pos, position; EA, effect allele; RA, reference allele; EAF, effect allele frequency; SE, standard error; ncRNA, non-coding RNA.

DISCUSSION

In the present study, we present the first GWAS of protein *O*-glycosylation, combined with clinical outcomes. We investigated the associations of apo-CIII proteoforms with lipid traits and prevalent type 2 diabetes, and performed GWASs on apo-CIII *O*-glycosylation to determine the direction of the associations. We found that the *GALNT2* locus was strongly associated with apo-CIII glycosylation and lipid traits. These findings were replicated in an independent population and then meta-analysed. The *GALNT2* locus linked apo-$CIII_{0a}$ and triglyceride levels to prevalent type 2 diabetes. To our knowledge, this is the first paper describing a genome-wide association with apo-CIII glycosylation. Moreover, a variant at the *IFT172*/*NRBP1* region, not previously linked to apo-CIII glycosylation, showed consistent associations with apo-CIII sialylation, and with triglyceride and total cholesterol levels.

We confirmed on a genome-wide significant scale that the *GALNT2*-gene plays a major role in the *O*-glycosylation of apo-CIII. The *GALNT2*-gene encodes the GalNAc-T2 enzyme that links N-acetylgalactosamine (GalNAc) to proteins, which is the essential first step in the addition of mucine-type *O*-linked glycans[192]. Apo-CIII is a selective GalNAc-T2 target: it is dependent of GalNAc-T2 for its glycosylation[193]. In our study, the *GALNT2* variants rs4846913-A and rs35498929-T were associated with a low proportion of non-glycosylated apo-$CIII_{0a}$, reflecting increased *O*-glycosylation of apo-CIII. The rs4846913 variant is located within a regulatory element that drives the expression of the *GALNT2*-gene in human hepatocytes[194], which is in line with our findings of increased apo-CIII glycosylation. *In vitro* studies suggest that the rs4846913-A allele increases binding with transcription factor CCAAT/enhancer binding protein beta (CEBPB), possibly leading to increased expression of *GALNT2*[194], while the C-allele has low functional activity[195].

Rs4846913-A also had a negative association with triglyceride levels, in line with findings in European and Asian consortia[196]. In contrast, Holleboom *et al.* found that rare *GALNT2* missense variants resulted in attenuated glycosylation of apo-CIII and better triglyceride clearance[135]. Holleboom *et al.* found altered apo-CIII sialylation

patterns in the carriers of a loss-of-function *GALNT2* variant[135]. They speculated that this was due to a smaller amount of glycosylated apo-CIII available for sialylation, increasing the number of disialylated species. In our study, sialylation within *O*-glycosylated species (apo-$CIII_{0c}$, apo-$CIII_1$, and apo-$CIII_2$) did not associate with the rs4846913 and rs35498929 *GALNT2* variants. Our genetic loci may not share the mechanism described by Holleboom et al. In our data, apo-$CIII_{0a}$ had a small positive association with triglyceride levels, which is in line with findings by Yassine et al.[58]. Apo-$CIII_{0a}$ was associated with higher HDL-c. Unfortunately, there is barely any literature about apo-$CIII_{0a}$ and HDL-c. Koska et al. found opposite directions of effect in their two study cohorts for the association of apo-$CIII_{0a}$ with HDL-c[57]. As demonstrated *in vitro*, the GalNAc-T2 enzyme, encoded by the *GALNT2*-gene, also glycosylates angiopoietin-like protein 3 (ANGPTL3), lecithin-cholesterol acyltransferase (LCAT), and phospholipid transfer protein (PLTP), which all play important roles in lipid metabolism, and especially triglyceride and HDL metabolism. Cholesteryl ester transfer protein (CETP) is another glycoprotein whose activity was directly associated with serum triglycerides and inversely with HDL-c[197]. We cannot exclude that *GALNT2* can affect lipid profiles through these or other proteins[198], this aspect should be elucidated in future studies.

Regarding sialylation, apo-$CIII_2$ and the apo-$CIII_1$/apo-$CIII_2$ ratio had strong negative and positive associations with triglyceride levels, respectively. Furthermore, rs67086575-G was associated with low apo-$CIII_2$, high apo-$CIII_1$/apo-$CIII_2$ ratio, and with high triglycerides in the meta-analyses. The variant also had a positive association with LDL-c, non-HDL-c and total cholesterol but this was confined to the DiaGene study. The association of this variant with triglycerides, LDL-c and total cholesterol is in agreement with findings in public GWAS data[189, 190]. Rs67086575 is located in the intronic region of the *IFT172*-gene. This gene encodes a peripheral subunit of the intraflagellar transport subcomplex IFT-B[199]. Mutations in this gene have been associated with ciliopathies, but so far not with glycosylation[199]. However, a link between primary cilia and obesity, insulin signalling, and type 2 diabetes have been described, which makes crosslinks to the lipoprotein metabolism not unlikely[200]. In blood cell lines, this SNP is associated with

increased expression of the *NRBP1*-gene[191]. This *NRBP1*-gene is suggested to play a role in subcellular trafficking between the endoplasmic reticulum and Golgi apparatus[201]. Since sialylation of *O*-glycans takes place at the Golgi-membrane[202], the link between rs67086575 and the *NRBP1*-gene might explain the association of this SNP with apo-CIII sialylation.

Apo-$CIII_2$ is associated with higher HDL-cholesterol, lower LDL-cholesterol and lower triglyceride levels, while apo-$CIII_{0c}$ and apo-$CIII_1$ showed opposite effects. Our findings are in accordance with findings in type 2 diabetes and prediabetes by Koska et al.[57]. Moreover, glycosylation of apo-CIII affects its inhibitory potential to LPL and influences the interaction of LDL with the vascular wall[135, 136]. From our data, we could speculate that lowering apo-$CIII_2$ or increasing apo-$CIII_1$ and apo-$CIII_{0a}$ might cause a shift towards a more atherogenic lipid profile. These glycoforms might be an interesting treatment target for the prevention of cardiovascular disease in the general population and in type 2 diabetes. Future large-scale research is needed to confirm the effects of apo-CIII proteoforms on lipid metabolism and cardiovascular disease risk, and to investigate the effect of medication on apo-CIII proteoform distributions, which is relevant for the development and effectivity of lipid lowering medications.

Hepatic receptors differentially clear sialylated apo-CIII glycoforms[59], therefore it has to be considered that changed proportions of certain apo-CIII proteoforms might not only be the result of aberrant posttranslational modification of apo-CIII, but might also be the result of accumulation due to dysfunctional clearance pathways. Nevertheless, our genetic findings pointed mainly towards the posttranslational modification of apo-CIII as a cause of aberrant glycosylation and sialylation. Larger GWASs might elucidate possible changes in clearance pathways of apo-CIII proteoforms.

This is the first GWAS of apo-CIII *O*-glycosylation, and therefore provides new insights into the genetic background of *O*-glycosylation. Current knowledge regarding the genetics of apo-CIII *O*-glycosylation is mainly based on studies addressing congenital disorders of glycosylation, where rare genetic variants have been shown to affect apo-

CIII *O*-glycosylation[153]. Findings in these disorders might not be applicable to the general population. Our findings had the same direction and magnitude of effect in people with and without diabetes, which suggests applicability to the general population. The DiaGene Study is the first cohort with detailed information on total plasma apo-CIII *O*-glycosylation on a large scale, using a high-throughput method to measure apo-CIII proteoforms. To date, the relations between lipids and apo-CIII glycoforms are only investigated in small cohorts. Our findings are in line with those studies, although non-glycosylated apo-$CIII_{0a,}$ has received little attention in the current literature.

There are no previous large-scale studies with data on apo-CIII *O*-glycosylation. Therefore, direct replication of our GWAS is not possible; this limits our power to detect small effect sizes. Nevertheless, we could only detect associations with substantial effect sizes, which are most relevant as a starting point. Notably, we did replicate the identified lead SNPs from the GWAS with plasma lipids in the independent Hoorn DCS cohort to improve power for effects on the clinical endpoints and decrease the risk of false-positive results. The Hoorn DCS cohort included patients with type 2 diabetes treated by their general practitioner, whereas the DiaGene study included patients with type 2 diabetes from all lines of care and people without diabetes. This could explain some of the differences between these cohorts. A few ssociations had opposing directions of effect in the DiaGene and Hoorn DCS cohorts. However, the confidence intervals were mostly overlapping, which doesn't allow definite conclusion on the the directions of effect might. Unfortunately, we do not have total apo-CIII plasma concentration data. The distribution of glycoforms in plasma is stable in variable apo-CIII concentrations in young, healthy men[203]. By adding the genetic basis to the analyses, we could offset this limitation for the most part. Both cohorts were mainly of Caucasian descent; therefore, we cannot generalise our findings to other ethnic groups.

In conclusion, variants of the *GALNT2*-gene affect apo-CIII *O*-glycosylation and lipid metabolism, influencing the risk of type 2 diabetes. The genetic variation in *GALNT2*, our main finding, is a well-known *O*-glycosylation enzyme. This suggests that the order of causality has the *O*-glycosylation at its basis, subsequently affecting apo-CIII function

and then altering the lipid profile, and not vice versa. Also, we propose the *NRBP1*-gene as a possible player in the sialylation of apolipoprotein-CIII and hypertriglyceridemia. Our results confirm a link between genomics, glycosylation, and lipid metabolism. This is a key step towards unravelling the regulation of serum and plasma glycoprotein *O*-glycosylation in health and disease. Glycophenotype characterisation alongside genetic variant identification should serve as relevant prognostic and predictive tools and should be considered for target identification of new pharmacological agents.

Acknowledgments

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 722095. This funding source had no involvement in the study design, collection, analysis, interpretation of data, writing or submitting the article.

The authors would like to thank the participants and staff of Diabetes Care System West-Friesland and the DiaGene Study for their cooperation and support.

CHAPTER 6

6. Apolipoprotein-CIII *O*-glycosylation is associated with micro- and macrovascular complications of type 2 diabetes

Annemieke Naber[1], Daniel Demus[2], Roderick Slieker[3,4], Simone Nicolardi[2], Joline W.J. Beulens[4,5,6], Petra J.M. Elders[7], Aloysius G. Lieverse[8], Eric J.G. Sijbrands[1], Leen M. 't Hart[3,4,5,9], Manfred Wuhrer[2], Mandy van Hoek[1]

[1] Department of Internal Medicine, Erasmus MC University Medical Center Rotterdam, Netherlands
[2] Center for Proteomics and Metabolomics, Leiden University Medical Center, the Netherlands
[3] Department of Cell and Chemical Biology, Leiden University Medical Center, the Netherlands
[4] Department of Epidemiology and Data Science, Amsterdam UMC, location Vrije Universiteit Amsterdam, the Netherlands
[5] Amsterdam Public Health, Amsterdam Cardiovascular Sciences, Amsterdam, The Netherlands
[6] Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands
[7] Department of General Practice, Amsterdam Public Health Institute, Amsterdam UMC, location VUmc, Amsterdam, the Netherlands
[8] Department of Internal Medicine, Maxima Medical Center, Eindhoven, the Netherlands
[9] Department of Biomedical Data Science, section Molecular Epidemiology, Leiden University Medical Center, the Netherlands

Reprinted and adapted: Naber, A.; Demus, D.; Slieker, R.C.; Nicolardi, S.; Beulens, J.W.J.; Elders, P.J.M.; Lieverse, A.G.; Sijbrands, E.J.G.; 't Hart, L.M.; Wuhrer, M.; et al. Apolipoprotein-CIII O-Glycosylation Is Associated with Micro- and Macrovascular Complications of Type 2 Diabetes. Int. J. Mol. Sci. 2024, 25, 5365. https://doi.org/10.3390/ijms25105365

ABSTRACT

Apolipoprotein-CIII (apo-CIII) inhibits the clearance of triglycerides from circulation and is associated with an increased risk of diabetes complications. It exists in four main proteoforms: O-glycosylated variants containing either zero, one, or two sialic acids and a non-glycosylated variant. O-glycosylation may affect the metabolic functions of apo-CIII. We investigated the associations of apo-CIII glycosylation in blood plasma, measured by mass spectrometry of the intact protein, and genetic variants with micro- and macrovascular complications (retinopathy, nephropathy, neuropathy, cardiovascular disease) of type 2 diabetes in a DiaGene study (n = 1571) and the Hoorn DCS cohort (n = 5409). Mono-sialylated apolipoprotein-CIII (apo-$CIII_1$) was associated with a reduced risk of retinopathy (β = −7.215, 95% CI −11.137 to −3.294) whereas disialylated apolipoprotein-CIII (apo-$CIII_2$) was associated with an increased risk (β = 5.309, 95% CI 2.279 to 8.339). A variant of the *GALNT2*-gene (rs4846913), previously linked to lower apo-$CIII_{0a}$, was associated with a decreased prevalence of retinopathy (OR = 0.739, 95% CI 0.575 to 0.951). Higher apo-$CIII_1$ levels were associated with neuropathy (β = 7.706, 95% CI 2.317 to 13.095) and lower apo-$CIII_{0a}$ with macrovascular complications (β = −9.195, 95% CI −15.847 to −2.543). In conclusion, apo-CIII glycosylation was associated with the prevalence of micro- and macrovascular complications of diabetes. Moreover, a variant in the GALNT2-gene was associated with apo-CIII glycosylation and retinopathy, suggesting a causal effect. The findings facilitate a molecular understanding of the pathophysiology of diabetes complications and warrant consideration of apo-CIII glycosylation as a potential target in the prevention of diabetes complications.

INTRODUCTION

Type 2 diabetes and its micro- and macrovascular complications pose worldwide problems in terms of morbidity, mortality, healthcare costs and low quality of life[204]. Despite treatment efforts, a substantial residual risk of these complications remains[205, 206]. Diabetic dyslipidaemia is one of the main risk factors for complications in type 2 diabetes and is characterised by a severely atherogenic lipid profile[207, 208]. Apolipoprotein-CIII (apo-CIII) levels have been linked to dyslipidaemia[209].

Apo-CIII is a protein involved in the metabolism of triglyceride-rich lipoproteins (TRLs)[123]. It has detrimental effects on the vascular wall by reducing the clearance of TRLs via lipoprotein lipase (LPL)-dependent and -independent pathways[160]. Apo-CIII enhances monocyte adhesion to the endothelium[210] and the binding of apoB-containing lipoproteins to vascular proteoglycans[122]. High levels of apo-CIII are found in individuals with diabetes mellitus[211, 212] and are related to reduced insulin sensitivity[58] and apoptosis of pancreatic beta-cells[172]. Apo-CIII levels have also been associated with diabetic retinopathy and nephropathy[213-215], and with increased cardiovascular disease risk in the general population and type 2 diabetes[127]. Genetic studies support the causality of this relationship with ischemic cardiovascular disease[170]. Consequently, it has been suggested that apo-CIII levels could be reduced by anti-apo-CIII antibody, antisense RNA, and silencing RNA therapies, to reduce vascular disease risk[130, 216, 217].

Posttranslational modifications, such as enzymatic glycosylation, can alter the function of apolipoproteins[59, 135, 136]. Apo-CIII exists in four main proteoforms: the native, non-glycosylated proteoform, and *O*-glycosylated proteoforms with either zero, one or two sialic acids[35, 58]. It has been recognised that glycosylation influences apo-CIII function and its relationship with lipid metabolism and cardiovascular disease[58, 59, 150]. Non- and monosialylated apo-CIII are associated with the formation of small-dense low-density lipoproteins (LDL)[218]. Altered apo-CIII glycosylation affects the inhibition of LPL by apo-CIII[135] and the interaction of LDL with the vascular wall[136]. Moreover, Apo-CIII

glycoforms are differentially cleared by hepatic receptors[59]. Different glycosylation patterns of apo-CIII have been observed in liver disease, metabolic syndrome, and the associated factors body weight and insulin sensitivity[10, 12]. Aberrant apo-CIII glycosylation patterns might increase the risk of complications in type 2 diabetes. These glycosylation patterns may have implications for the efficacy of preventive therapies addressing diabetes complications and in particular for lipid lowering agents in type 2 diabetes.

In the present study, we investigated the relation of apo-CIII glycosylation with the prevalence and incidence of diabetic retinopathy, nephropathy, neuropathy, and macrovascular complications. The direction of these relationships was assessed with recently identified genetic variants associated with apo-CIII glycosylation. We used data from the DiaGene cohort, a prospective study of type 2 diabetes conducted in Eindhoven, the Netherlands[141]. Genetic associations were replicated and meta-analysed in an independent cohort with type 2 diabetes patients in West-Friesland, the Netherlands: the Hoorn Diabetes Care System (Hoorn DCS)[180].

RESEARCH DESIGN AND METHODS

Study design

Plasma samples from the DiaGene study were used. In short, the DiaGene study is an all case-control study addressing all lines of healthcare with prospective follow-up in the cities of Eindhoven and Veldhoven, the Netherlands. The DiaGene study sought to include all type 2 diabetes patients in hospital and primary care at a given time point in the past. A prospective follow-up was performed from that moment onwards. Inclusion took place between 2006 and 2011, with a median follow-up duration of 7.8 years. This study has been described in detail elsewhere[141]. The current study was restricted to cases with type 2 diabetes for which apo-CIII glycosylation measurements were available (n=1,571).

Data from the Hoorn DCS (n=5,409) were used for replication of genetic associations with complications of type 2 diabetes. This cohort has been described in detail elsewhere[180]. In short, biobanking materials and data on annual examinations for micro- and macrovascular complications have been collected from primary care type 2 diabetes patients, with prospective follow-up. The Hoorn DCS study included newly diagnosed type 2 diabetes patients. Afterwards, follow-up was performed. Inclusion took place between 1998 and 2014 in the region of West Friesland in the Netherlands. The median follow-up duration was 9.0 years.

All participants gave their written informed consent. Both studies were approved by the Medical Ethics Committees of the involved hospitals in compliance with the Declaration of Helsinki principles (DiaGene MEC-2004-230, Hoorn DCS 2007/57).

Definitions

Type 2 diabetes was defined in accordance with the American Diabetes Association and the WHO guidelines[219, 220] for both the DiaGene and the Hoorn DCS cohorts[141, 180]. People with other types of diabetes were excluded from both cohorts.

Retinopathy, nephropathy, and neuropathy were considered microvascular complications. Diabetic retinopathy was diagnosed using fundus photography. Diabetic

nephropathy was defined as albumin/creatinine ratio (ACR) ≥ 2.5 for men or ≥ 3.5 for women present at two of three consecutive measurements, or when ACR ≥ 12.5 for men or ≥ 17.5 for women was present at one measurement. Diabetic neuropathy was defined by a podiatrist, neurologist or the patient's treating physician. Only hospital treated participants of the DiaGene study had information about diabetic neuropathy (n=611). Macrovascular complications in the DiaGene study comprised ischemic heart disease (myocardial infarction, percutaneous coronary intervention or coronary artery bypass graft), cerebrovascular accident, transient ischemic attack and peripheral arterial disease. In the DiaGene, information on cardiovascular disease for hospital-treated patients was retrieved from the medical records, and for primary care-treated patients from self-reporting[141]. In the Hoorn DCS, cardiovascular events were reported during the annual visits and verified against the medical records from the regional hospital and general practitioners. More detailed information on the definitions and data collection of micro- and macrovascular complications of diabetes has been described previously[180].

Apo-CIII glycosylation analysis

Ultrahigh-resolution matrix-assisted laser desorption/ionization Fourier transform ion cyclotron resonance mass spectrometry (MALDI FT-ICR MS) method, described elsewhere[179], was applied to assess the relative levels of apo-CIII proteoforms in blood plasma samples of the DiaGene cohort. Apo-CIII exists in four main proteoforms: glycosylated variants containing a mucin-type core-1 *O*-glycan with either zero, one or two sialic acids (apo-$CIII_{0c}$, apo-$CIII_1$ and apo-$CIII_2$ respectively) and a non-glycosylated variant (apo-$CIII_{0a}$) (**Figure 1**)[35, 58]. The sum of apo-$CIII_{0c}$, apo-$CIII_1$, and apo-$CIII_2$ was set to 1.0 to obtain the proportion of these three glycoforms within all glycosylated species of apo-CIII. These normalised relative peak intensities of the glycosylated variants apo-$CIII_{0c}$, apo-$CIII_1$, and apo-$CIII_2$ were used to assess associations with apo-CIII sialylation status. The relative peak intensity of apo-$CIII_{0a}$, normalised to the sum of all four proteoforms (apo-$CIII_{0a}$, apo-$CIII_{0c}$, apo-$CIII_1$, and apo-$CIII_2$), was used to assess associations with apo-CIII glycosylation status. We used beta to present the outcome

measures of our regression models on glycosylation traits to ease the interpretation. Quality control of mass spectrometry (MS) data was performed as described elsewhere[179]. Samples not meeting the acceptable quality parameters were excluded from the subsequent analysis. Apo-CIII glycosylation data was only available in the DiaGene study.

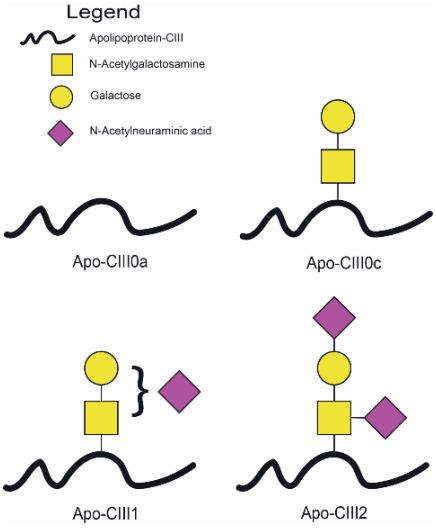


Figure 1. Apo-CIII glycoforms. Apo-CIII has four main proteoforms: apo-$CIII_{0a}$ is the non-glycosylated form; apo-$CIII_{0c}$ is glycosylated, without sialic acid; and apo-$CIII_1$ and apo-$CIII_2$ are glycosylated with one or two sialic acids respectively. Black symbol = apo-CIII protein; yellow square = *N*-acetylgalactosamine; yellow circle = galactose; purple diamond = *N*-acetylneuraminic acid (sialic acid).

Experimental design and statistical analysis

A total of 1571 DiaGene samples passed apo-CIII glycosylation data quality control and contained sufficient clinical information for the analysis. Missing data on covariates (duration of diabetes, HbA1c, use of lipid-lowering medication, and use of insulin) within the DiaGene study were imputed using multiple imputations by predictive mean matching in SPSS. The maximum count of imputations per variable was 105 in DiaGene; all imputed variables had <7% missing values. Covariates within the DCS Hoorn had <2% missing values per covariate, and therefore were not imputed (**Supplementary Table S1**).

Table S1. Number of missings per variable.

	DiaGene		DCS Hoorn	
Variable	Number	Percentage	Number	Percentage
Sex	0	0.00	0	0.00
Age	0	0.00	0	0.00
Duration of diabetes	105	6.7	37	0.68
BMI	107	6.8	39	0.72
HbA1c	78	5.0	107	1.98
Systolic blood pressure	105	6.7	76	1.41
Diastolic blood pressure	104	6.6	76	1.41
Mean arterial pressure	105	6.7	76	1.41
Creatinine	102	6.5	113	2.09
HDL-cholesterol	80	5.1	103	1.90
non-HDL-cholesterol	80	5.1	104	1.92
LDL-cholesterol	102	6.5	194	3.59
Total cholesterol	77	4.9	98	1.81
Triglycerides	79	5.0	100	1.85
Lipid lowering medication	97	6.2	0	0.00
Insulin or analogues	97	6.2	0	0.00
Smoking	149	9.5	183	3.38

Associations of apo-CIII glycosylation with micro- and macrovascular complications at baseline and follow-up were investigated using logistic regression and Cox proportional hazards models, respectively. Prospective analyses were performed after excluding prevalent complications. Analyses were performed for retinopathy, nephropathy, neuropathy, and macrovascular complications separately.

Two models were applied for all analyses. Model 1 was adjusted for age and sex. Model 2 was adjusted for age, sex, haemoglobin A1c (HbA1c), and duration of diabetes. The outcomes of Model 1 reflect the broad differences between diabetes patients with and without one of the complications. In contrast, Model 2 reflects the differences not mediated through the duration of type 2 diabetes or its regulation, with HbA1c as a proxy.

We performed two-sided *t*-tests to analyse the associations of insulin use and lipid-lowering therapy with apo-CIII glycosylation. Subsequently, we performed two sensitivity analyses: adding insulin use to Model 1, and the use of lipid-lowering therapy (fibrates and statins) to Model 2.

In our previous genome-wide association study (GWAS) of the apo-CIII glycosylation[221], we identified genetic variants associated with apo-CIII *O*-glycosylation. We selected two loci from this GWAS, with previous links to triglyceride levels. We analysed the associations of these genetic variants with micro- and macrovascular complications, applying the same models as described above. These analyses were replicated in the Hoorn DCS cohort and meta-analysed using a random effects model for the two cohorts using the package 'meta' version 7.0-0[222]. Heterogeneity between cohorts was also evaluated using I^2 values (**Supplementary Table S2**)[223].

Statistical analyses within the DiaGene study were performed using SPSS version 25. Analyses of the Hoorn Diabetes Care System (DCS) cohort and the meta-analysis were performed using R version 4.0.5. For the analysis with apo-CIII glycosylation, the Bonferroni corrected *p*-value for significance was calculated as 0.013 (0.05/4) based on the four main apo-CIII proteoforms. For the analyses of the four genetic variants and complications of diabetes, we used a Bonferroni corrected *p*-value for a significance of 0.013 (0.05/4). Significance is given after Bonferroni correction unless stated otherwise.

Table S2. Measures of heterogeniety.

Retinopathy

		Prevalence		Incidence	
		Model 1	Model 2	Model 1	Model 2
uniqID	rsID (dbSNP build 146)	I2	I2	I2	I2
1:230294715:A:C	rs4846913	0	0.459	0	0.511
1:230286016:C:T	rs35498929	0	0	0.756	0.652
1:230416320:C:T	rs3213497	0.104	0.496	0	0
2:27686480:A:G	rs67086575	0	0	0	0.116

Nephropathy

		Prevalence		Incidence	
		Model 1	Model 2	Model 1	Model 2
uniqID	rsID (dbSNP build 146)	I2	I2	I2	I2
1:230294715:A:C	rs4846913	0	0	0.270	0.044
1:230286016:C:T	rs35498929	0	0	0	0
1:230416320:C:T	rs3213497	0	0	0	0
2:27686480:A:G	rs67086575	0.605	0.495	0	0

Macrovascular complications

		Prevalence		Indicence	
		Model 1	Model 2	Model 1	Model 2
uniqID	rsID (dbSNP build 146)	I2	I2	I2	I2
1:230294715:A:C	rs4846913	0	0	0	0
1:230286016:C:T	rs35498929	0.390	0.415	0	0
1:230416320:C:T	rs3213497	0.132	0	0.568	0.681
2:27686480:A:G	rs67086575	0	0	0.640	0.406

RESULTS

Cohort characteristics

The characteristics of cases with type 2 diabetes of both cohorts are presented in Table 1; 45.8% and 44.6% were female, and the mean age was 65.1 (SD 10.6) and 61.1 (SD 11) years in the DiaGene and Hoorn DCS, respectively. Compared with the Hoorn DCS, individuals from the DiaGene study had a longer duration of type 2 diabetes (8.0 vs. 0.6 years). At inclusion, the prevalence of complications was higher in the DiaGene than in the Hoorn DCS for retinopathy (17.4 vs. 3.6%), nephropathy (22.1 vs. 9.2%), and macrovascular complications (40.4 vs. 2.1%). In the DiaGene study, 30.6% of the participants had neuropathy at inclusion, but the Hoorn DCS did not collect data on neuropathy. The incidence of complications during follow-up was comparable between the DiaGene and the Hoorn DCS (**Table 1**).

Table 1. General characteristics of the study populations

	DiaGene (n=1,571)	**Hoorn DCS (n=5,409)**
Female sex, n (%)	719 (45.8)	2,414 (44.6)
Age, year, mean (±SD)	65.1 (10.6)	61.1 (11)
Duration of diabetes, year, median (IQR)	8.0 (10.6)	0.6 (2.7)
BMI, kg/m2, median (IQR)	30.0 (6.3)	29.4 (6.5)
HbA1c, %, median (IQR)	6.8 (1.3)	6.7 (1.4)
HbA1c, mmol/mol, median (IQR)	50.8 (14.2)	49.7 (15.8)
Systolic blood pressure, mmHg, mean (±SD)	142.0 (19.0)	142.5 (20)
Diastolic blood pressure, mmHg, mean (±SD)	77.5 (9.9)	80.8 (10)
Mean arterial pressure, mmHg, mean (±SD)	99.0 (11.0)	101.4 (12)
Creatinine, µmol/l, median (IQR)	77.0 (25.0)	79.9 (21.7)
HDL-cholesterol, mmol/L, median (IQR)	1.1 (0.4)	1.2 (0.4)
Non-HDL-cholesterol, mmol/L, median (IQR)	3.0 (1.1)	3.8 (1.5)
LDL-cholesterol, mmol/L, mean (±SD)	2.5 (0.8)	2.9 (1.5)
Total Cholesterol, mmol/L, median (IQR)	4.2 (1.1)	5.0 (1.6)
Triglycerides, mmol/L, median (IQR)	1.4 (1.0)	1.7 (1.1)
Medication use, n (%)		
Statins or fibrates	1020 (69.2)	2276 (42.1)
Insulin or analogues	463 (31.4)	447 (8.3)
Smoking, n (%)		
Never	361 (25.4)	1,667 (31.9)
Former	804 (56.5)	2,429 (46.5)
Current	257 (18.1)	1,130 (21.6)
Complications, n-case/n-total (%)		
Prevalent Retinopathy	257/1,478 (17.4)	194/5,321 (3.6)
Incident Retinopathy	175/1,168 (15.0)	824/5,127 (16.1)
Prevalent Nephropathy	309/1,401 (22.1)	499/5,407 (9.2)
Incident Nephropathy	206/1,058 (19.5)	1009/4,908 (20.6)
Prevalent Neuropathy	187/611 (30.6)	NA

Incident Neuropathy	192/400 (48.0)	NA
Prevalent macrovascular disease	536/1,328 (40.4)	111/5,396 (2.1)
Incident macrovascular disease	86/913 (9.4)	438/5,285 (8.3)

BMI: body mass index; HDL: high-density lipoprotein; LDL: low density lipoprotein. Continuous data are presented as mean (and standard deviation) or median (and interquartile range) for normal and non-normal distributions respectively. The distribution of the clinical variables was considered normal when Skewness and Kurtosis were within the range of −1 to +1.

Associations of apo-CIII glycosylation with micro- and macrovascular complications in the DiaGene Study

Apo-$CIII_1$ and apo-$CIII_2$ levels showed strong associations with the prevalence (**Table 2**) and incidence of retinopathy (**Table 3**). In Model 1, apo-$CIII_1$ was associated with a decreased prevalence of retinopathy (β = −7.215, 95% CI −11.137 to −3.294), while apo-$CIII_2$ was associated with an increased prevalence (β = 5.309, 95% CI 2.279 to 8.339). These associations lost significance in Model 2, and the effect size was slightly reduced but showed the same trend. The association of apo-$CIII_{0a}$ with an increased prevalence of retinopathy reached suggestive significance in Model 2 (β = 9.968, 95% CI 1.437 to 18.499), after adjustment for duration of diabetes and HbA1c. Apo-$CIII_2$ was significantly associated with incident retinopathy in Model 1 (β = 4.484, 95% CI 1.158 to 7.810). The associations of apo-$CIII_1$ and apo-$CIII_2$ with prevalent and incident retinopathy had consistent directions of effect.

For nephropathy, apo-$CIII_1$ was associated with a decreased prevalence (β = −3.945, 95% CI 7.716 to −0.174), but did not reach Bonferroni-corrected significance (p = 0.040) (**Table 2**). Adding covariates in Model 2 did not alter the direction of the association. Also, the direction and magnitude of the effect for the association of apo-$CIII_1$ and incident nephropathy were similar to the association found at baseline (**Table 3**).

Apo-$CIII_1$ and apo-$CIII_2$ levels were associated with neuropathy (**Table 2**). Apo-$CIII_1$ was significantly associated with an increased prevalence of neuropathy (β = 7.706, 95% CI 2.317 to 13.095), while apo-$CIII_2$ was associated with a decreased prevalence of neuropathy without reaching Bonferroni-corrected significance (β = −4.968, 95% CI −9.065 to −0.871, p = 0.017). The direction of effects and their significance were similar in Model 2. The analysis for incident neuropathy did not reach significance but showed a similar direction of effects (**Table 3**).

Apo-$CIII_{0a}$ levels were associated with macrovascular complications of diabetes (**Table 2**). This glycoform had a significant negative association with the prevalence of macrovascular complications in Model 1 (β = −9.195, 95% CI −15.847 to −2.543), which remained significant in Model 2 and the sensitivity analysis for insulin use (**Table 2 and Supplementary Table S3**). This association lost significance in the analysis of the incidence of macrovascular complications (**Table 3**). Nevertheless, the effect size and trend were similar.

The use of lipid-reducing agents was associated with lower relative levels of apo-CIII0a and insulin use was associated with lower apo-$CIII_1$ and higher apo-$CIII_2$ (**Supplementary Table S4**). Most associations persisted after sensitivity analysis for the use of lipid-lowering medication and insulin (**Supplementary Table S3**).

Table 2. Associations of apo-CIII proteoforms with prevalent complications of type 2 diabetes in the DiaGene study.

Retinopathy						
	Model 1			**Model 2**		
Proteoform	**Beta**	**95% CI**	***p*-Value**	**Beta**	**95% CI**	***p*-Value**
Apo-$CIII_{0a}$	4.053	−3.502 to 11.609	0.293	9.968	1.437 to 18.499	**0.022**
Apo-$CIII_{0c}$	−3.930	−9.252 to 1.392	0.148	−3.372	−9.213 to 2.468	0.258
Apo-$CIII_1$	−7.215	−11.137 to −3.294	**0.0003**	−4.121	−8.543 to 0.301	0.068
Apo-$CIII_2$	5.309	2.279 to 8.339	**0.0006**	3.379	−0.024 to 6.783	0.052
Nephropathy						
Apo-$CIII_{0a}$	0.486	−6.938 to 7.910	0.898	2.403	−5.132 to 9.938	0.532
Apo-$CIII_{0c}$	0.363	−4.519 to 5.244	0.884	0.890	−4.082 to 5.861	0.726
Apo-$CIII_1$	−3.945	−7.716 to −0.174	**0.040**	−2.468	−6.338 to 1.402	0.211
Apo-$CIII_2$	2.377	−0.512 to 5.266	0.107	1.386	−1.589 to 4.361	0.361
Neuropathy						
Apo-$CIII_{0a}$	5.505	−4.277 to 15.287	0.270	5.871	−3.979 to 15.720	0.243
Apo-$CIII_{0c}$	1.064	−5.392 to 7.520	0.747	0.951	−5.545 to 7.447	0.774
Apo-$CIII_1$	7.706	2.317 to 13.095	**0.005**	8.000	2.558 to 13.441	**0.004**
Apo-$CIII_2$	−4.968	−9.065 to −0.871	**0.017**	−5.116	−9.260 to −0.972	**0.016**
Macrovascular disease						
Apo-$CIII_{0a}$	−9.195	−15.847 to −2.543	**0.007**	−8.795	−15.471 to −2.118	**0.010**
Apo-$CIII_{0c}$	0.635	−3.615 to 4.885	0.770	0.880	−3.392 to 5.152	0.686
Apo-$CIII_1$	−0.123	−3.429 to 3.184	0.942	0.301	−3.042 to 3.644	0.860
Apo-$CIII_2$	−0.077	−2.622 to 2.468	0.953	−0.401	−2.979 to 2.176	0.760

Model 1: adjusted for age and sex; Model 2: adjusted for age, sex, duration of diabetes, and HbA1c. Apo-$CIII_{0a}$ normalised to all four proteoforms: apo-$CIII_{0a}$, apo-$CIII_{0c}$, apo-$CIII_1$, and apo-$CIII_2$. The sum of the glycoforms Apo-$CIII_{0c}$, Apo-$CIII_1$, and Apo-$CIII_2$ was set to 1.0. Beta represents the change of prevalence of the respective complication per increase of 1 standard deviation of the relative peak intensity of the proteoform of apo-CIII. Bold for $p < 0.05$, bold and underlined for $p < 0.013$.

Table 3. Associations of apo-CIII proteoforms with incident complications of type 2 diabetes in the DiaGene study.

	Retinopathy					
	Model 1			Model 2		
Proteoform	**Beta**	**95% CI**	***p*-Value**	**Beta**	**95% CI**	***p*-Value**
Apo-$CIII_{0a}$	−4.554	−13.457 to 4.348	0.316	3.171	−5.753 to 12.094	0.486
Apo-$CIII_{0c}$	−5.834	−11.866 to 0.198	0.058	−4.784	−10.840 to 1.271	0.122
Apo-$CIII_1$	−4.958	−9.268 to −0.649	**0.024**	−2.037	−6.521 to 2.447	0.373
Apo-$CIII_2$	4.484	1.158 to 7.810	**0.008**	2.522	−0.962 to 6.006	0.156
	Nephropathy					
Apo-$CIII_{0a}$	−5.189	−13.555 to 3.178	0.224	−4.794	−13.196 to 3.608	0.263
Apo-$CIII_{0c}$	−0.130	−5.465 to 5.204	0.962	−0.488	−5.831 to 4.855	0.858
Apo-$CIII_1$	−2.852	−6.933 to 1.229	0.171	−2.761	−6.936 to 1.413	0.195
Apo-$CIII_2$	1.366	−1.863 to 4.595	0.407	1.400	−1.877 to 4.677	0.402
	Neuropathy					
Apo-$CIII_{0a}$	−5.944	−14.462 to 2.573	0.171	−5.555	−14.206 to 3.097	0.208
Apo-$CIII_{0c}$	3.525	−1.469 to 8.519	0.166	2.213	−2.896 to 7.323	0.396
Apo-$CIII_1$	0.647	−3.527 to 4.821	0.761	0.398	−3.896 to 4.691	0.856
Apo-$CIII_2$	−1.981	−5.206 to 1.245	0.229	−1.308	−4.601 to 1.984	0.436
	Macrovascular Disease					
Apo-$CIII_{0a}$	−6.948	−19.786 to 5.890	0.289	−3.852	−16.455 to 8.751	0.549
Apo-$CIII_{0c}$	0.564	−7.660 to 8.788	0.893	0.992	−6.978 to 8.962	0.807
Apo-$CIII_1$	−2.327	−8.512 to 3.857	0.461	0.244	−6.016 to 6.504	0.939
Apo-$CIII_2$	0.639	−4.317 to 5.595	0.800	−1.096	−6.055 to 3.864	0.665

Model 1: adjusted for age and sex; Model 2: adjusted for age, sex, duration of diabetes, and HbA1c. Apo-$CIII_{0a}$ normalised to all four proteoforms: apo-$CIII_{0a}$, apo-$CIII_{0c}$, apo-$CIII_1$, and apo-$CIII_2$. The sum of the glycoforms Apo-$CIII_{0c}$, Apo-$CIII_1$, and Apo-$CIII_2$ was set to 1.0. Beta represents the change of incidence of the respective complication per increase of 1 standard deviation of the relative peak intensity of the proteoform of apo-CIII. Bold for $p < 0.05$, bold and underlined for $p < 0.013$.

Table S3. Association of apo-CIII proteoforms with prevalent and incident complications of type 2 diabetes, sensitivity analyses in the DiaGene cohort.

Prevalence

Retinopathy	Sens. Lipid Meds				Sens. Insulin use			
		95% C.I. for beta				95% C.I. for beta		
Proteoform	Beta	lower	upper	p-value	Beta	lower	upper	p-value
Apo-$CIII_{0a}$	8.935	0.314	17.556	**0.042**	4.115	-3.784	12.013	0.307
Apo-$CIII_{0c}$	-3.202	-9.049	2.644	0.283	-3.901	-9.431	1.629	0.167
Apo-$CIII_1$	-4.355	-8.771	0.061	0.053	-5.488	-9.626	-1.350	**0.009**
Apo-$CIII_2$	3.460	0.065	6.855	**0.046**	4.329	1.125	7.533	**0.008**

Prevalence

Nephropathy	Sens. Lipid Meds				Sens. Insulin use			
		95% C.I. for beta				95% C.I. for beta		
Proteoform	Beta	lower	upper	p-value	Beta	lower	upper	p-value
Apo-$CIII_{0a}$	2.504	-5.094	10.103	0.518	0.332	-7.121	7.784	0.931
Apo-$CIII_{0c}$	0.883	-4.088	5.854	0.728	0.378	-4.528	5.283	0.880
Apo-$CIII_1$	-2.460	-6.334	1.414	0.213	-3.323	-7.135	0.490	0.088
Apo-$CIII_2$	1.383	-1.594	4.359	0.363	2.015	-0.903	4.933	0.176

Prevalence

Neuropathy	Sens. Lipid Meds				Sens. Insulin use			
		95% C.I. for beta				95% C.I. for beta		
Proteoform	Beta	lower	upper	p-value	Beta	lower	upper	p-value
Apo-$CIII_{0a}$	5.602	-4.342	15.545	0.270	5.255	-4.579	15.089	0.295
Apo-$CIII_{0c}$	0.924	-5.572	7.421	0.780	1.061	-5.414	7.537	0.748
Apo-$CIII_1$	7.933	2.487	13.379	**0.004**	8.190	2.741	13.640	**0.003**
Apo-$CIII_2$	-5.061	-9.209	-0.914	**0.017**	-5.221	-9.349	-1.093	**0.013**

Prevalence

Cardiovascular disease	Sens. Lipid Meds				Sens. Insulin use			
		95% C.I. for beta				95% C.I. for beta		
Proteoform	Beta	lower	upper	p-value	Beta	lower	upper	p-value
Apo-$CIII_{0a}$	-6.565	-13.368	0.239	0.059	-9.216	-15.874	-2.558	**0.007**
Apo-$CIII_{0c}$	0.805	-3.525	5.135	0.716	0.715	-3.546	4.976	0.742
Apo-$CIII_1$	0.778	-2.636	4.192	0.655	0.184	-3.142	3.509	0.914
Apo-$CIII_2$	-0.665	-3.301	1.971	0.621	-0.282	-2.839	2.276	0.829

	Incidence							
Retinopathy	Sens. Lipid Meds				Sens. Insulin use			
		95% C.I. for beta				95% C.I. for beta		
Proteoform	Beta	lower	upper	p-value	Beta	lower	upper	p-value
Apo-CIII$_{0a}$	3.219	-5.728	12.165	0.481	-3.151	-11.876	5.574	0.479
Apo-CIII$_{0c}$	-4.788	-10.849	1.273	0.122	-5.500	-11.426	0.427	0.069
Apo-CIII$_1$	-2.041	-6.529	2.447	0.373	-2.076	-6.386	2.233	0.345
Apo-CIII$_2$	2.527	-0.964	6.018	0.156	2.820	-0.516	6.157	0.098

	Incidence							
Nephropathy	Sens. Lipid Meds				Sens. Insulin use			
		95% C.I. for beta				95% C.I. for beta		
Proteoform	Beta	lower	upper	p-value	Beta	lower	upper	p-value
Apo-CIII$_{0a}$	-3.998	-12.547	4.550	0.359	-5.179	-13.549	3.191	0.225
Apo-CIII$_{0c}$	-0.669	-5.998	4.659	0.806	-0.131	-5.466	5.204	0.962
Apo-CIII$_1$	-2.566	-6.790	1.658	0.234	-2.949	-7.080	1.183	0.162
Apo-CIII$_2$	1.324	-1.971	4.619	0.431	1.402	-1.847	4.651	0.398

	Incidence							
Neuropathy	Sens. Lipid Meds				Sens. Insulin use			
		95% C.I. for beta				95% C.I. for beta		
Proteoform	Beta	lower	upper	p-value	Beta	lower	upper	p-value
Apo-CIII$_{0a}$	-5.910	-14.638	2.819	0.185	-6.023	-14.551	2.506	0.166
Apo-CIII$_{0c}$	2.161	-2.949	7.271	0.407	3.527	-1.457	8.512	0.165
Apo-CIII$_1$	0.297	-4.001	4.595	0.892	0.637	-3.542	4.817	0.765
Apo-CIII$_2$	-1.236	-4.539	2.066	0.463	-1.981	-5.202	1.240	0.228

	Incidence							
Cardiovascular disease	Sens. Lipid Meds				Sens. Insulin use			
		95% C.I. for beta				95% C.I. for beta		
Proteoform	Beta	lower	upper	p-value	Beta	lower	upper	p-value
Apo-CIII$_{0a}$	-3.356	-16.092	9.379	0.605	-6.659	-19.420	6.102	0.306
Apo-CIII$_{0c}$	0.903	-7.034	8.840	0.824	0.912	-7.350	9.174	0.829
Apo-CIII$_1$	0.335	-5.964	6.633	0.917	-1.252	-7.459	4.956	0.693
Apo-CIII$_2$	-1.122	-6.094	3.849	0.658	-0.155	-5.172	4.862	0.952

Bold for p<0.05; bold and underlined for p<0.05/4=0.013

Table S4. Association of Lipid lowering medication and insulin use with apo-CIII proteoforms in the DiaGene cohort

	Lipid Meds					Insulin use				
	Users		Non-users		p-value	Users		Non-users		p-value
Proteoform	Mean	SD	Mean	SD		Mean	SD	Mean	SD	
Apo-$CIII_{0a}$	0.035	0.017	0.039	0.019	**0.0003**	0.036	0.018	0.036	0.017	0.552
Apo-$CIII_{0c}$	0.114	0.027	0.114	0.026	0.742	0.113	0.026	0.115	0.027	0.261
Apo-$CIII_1$	0.670	0.035	0.674	0.037	0.058	0.667	0.036	0.673	0.035	**0.004**
Apo-$CIII_2$	0.215	0.045	0.211	0.049	0.184	0.219	0.045	0.211	0.046	**0.005**

Bold for $p<0.05$; bold and underlined for $p<0.05/4=0.013$. Mean (and standard deviation). P-values are for use versus non-use of the mentioned medication.

Apo-CIII glycosylation-associated genetic variants and complications of diabetes in the DiaGene and Hoorn DCS studies, a meta-analysis

The genetic variant rs4846913-A located in the *GALNT2* gene, previously associated with decreased apo-$CIII_{0a}$[221]. was found to be negatively associated with prevalent retinopathy in Model 2 in the DiaGene study (OR = 0.739, 95% CI 0.575 to 0.951). However, it did not reach the Bonferroni significance level (**Table 4, Table 5, Figure 1**). This association remained in the sensitivity analysis for the use of lipid-lowering medications (OR = 0.722, 95% CI 0.559 to 0.932) and lost its significance after the sensitivity analysis for insulin use (**Supplementary Table S5**). These associations had the same direction of effect in the Hoorn DCS cohort.

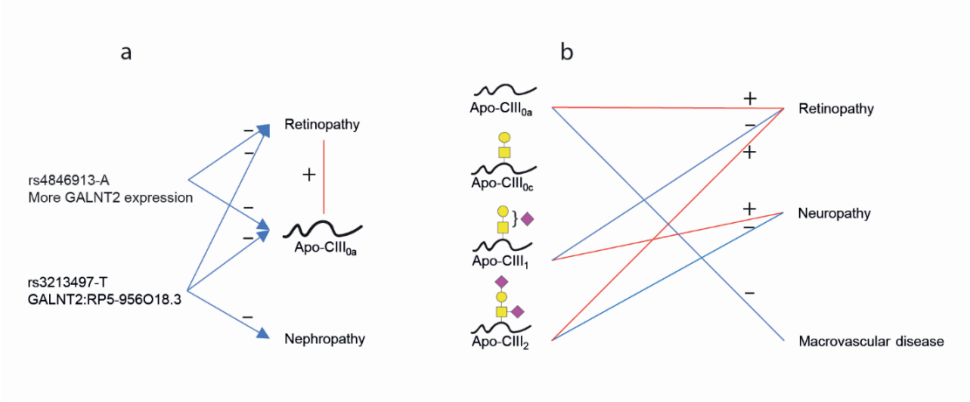


Figure 2. Summary of associations of genetic variants, apo-CIII proteoforms and complications of diabetes. a) Rs4846913-A is associated with lower apo-$CIII_{0a}$ and lower prevalence of retinopathy. Rs3213497-T is associated with lower apo-$CIII_{0a}$ and decreased incidence of retinopathy and prevalence of nephropathy. Apo-$CIII_{0a}$ is associated with higher prevalence of retinopathy. b) The four proteoforms of apo-CIII are associated with micro- and macrovascular complications of diabetes. Blue: negative associations, red: positive associations.

Table 4. Associations of genetic variants with prevalent complications of type 2 diabetes, Model 1 (adjusted for age and sex).

Retinopathy					DiaGene		Hoorn DCS			Meta-Analysis	
Locus	**rsID**	**EA**	**RA**	**OR**	**95% CI**	***p*-Value**	**OR**	**95% CI**	***p*-Value**	**OR**	***p*-Value**
GALNT2	rs4846913	A	C	0.869	0.696 to 1.086	0.218	0.935	0.763 to 1.147	0.521	0.905	0.192
GALNT2	rs35498929	T	C	0.884	0.637 to 1.227	0.461	0.897	0.651 to 1.236	0.508	0.891	0.324
GALNT2:RP5-956O18.3	rs3213497	T	C	0.991	0.716 to 1.372	0.958	0.774	0.559 to 1.071	0.122	0.875	0.283
IFT172/NRBP1	rs67086575	G	A	0.819	0.594 to 1.128	0.221	0.929	0.701 to 1.231	0.606	0.879	0.231
Nephropathy					**DiaGene**		**Hoorn DCS**			**Meta-Analysis**	
Locus	**rsID**	**EA**	**RA**	**OR**	**95% CI**	***p*-Value**	**OR**	**95% CI**	***p*-Value**	**OR**	***p*-Value**
GALNT2	rs4846913	A	C	0.912	0.988 to 0.795	0.912	1.033	0.904 to 1.179	0.637	1.020	0.730
GALNT2	rs35498929	T	C	0.999	0.733 to 1.360	0.993	0.903	0.734 to 1.110	0.333	0.932	0.417
GALNT2:RP5-956O18.3	rs3213497	T	C	0.853	0.619 to 1.175	0.331	0.819	0.667 to 1.005	0.056	0.829	**0.033**
IFT172/NRBP1	rs67086575	G	A	0.282	0.848 to 0.628	0.282	1.123	0.945 to 1.335	0.189	1.003	0.980
Neuropathy					**DiaGene**		**Hoorn DCS**			**Meta-Analysis**	
Locus	**rsID**	**EA**	**RA**	**OR**	**95% CI**	***p*-Value**	**OR**	**95% CI**	***p*-Value**	**OR**	***p*-Value**
GALNT2	rs4846913	A	C	1.270	0.948 to 1.700	0.109	NA	NA	NA	NA	NA
GALNT2	rs35498929	T	C	1.350	0.895 to 2.034	0.152	NA	NA	NA	NA	NA
GALNT2:RP5-956O18.3	rs3213497	T	C	1.161	0.765 to 1.761	0.483	NA	NA	NA	NA	NA

IFT172/NRBP1	rs67086575	G	A	0.989	0.664 to 1.472	0.956	NA	NA	NA	NA	NA
Macrovascular Complications					**DiaGene**		**Hoorn DCS**			**Meta-Analysis**	
Locus	**rsID**	**EA**	**RA**	**OR**	**95% CI**	***p*-Value**	**OR**	**95% CI**	***p*-Value**	**OR**	***p*-Value**
GALNT2	rs4846913	A	C	1.103	0.922 to 1.320	0.283	1.056	0.806 to 1.385	0.691	1.014	0.869
GALNT2	rs35498929	T	C	0.917	0.711 to 1.184	0.507	1.236	0.846 to 1.806	0.273	1.029	0.842
GALNT2:RP5-956O18.3	rs3213497	T	C	0.983	0.755 to 1.280	0.900	1.256	0.875 to 1.803	0.216	1.076	0.534
IFT172/NRBP1	rs67086575	G	A	0.937	0.736 to 1.194	0.601	0.840	0.572 to 1.234	0.375	0.909	0.358

Beta represents the change of prevalence of the respective complication per increase of the effect allele. Bold for $p < 0.05$. EA, effect allele; RA, reference allele; NA, not available.

6

Table 5. Associations of genetic variants with prevalent complications of type 2 diabetes, Model 2 (adjusted for age, sex, duration of diabetes, and HbA1c).

Retinopathy					**DiaGene**		**Hoorn DCS**			**Meta-Analysis**	
Locus	**rsID**	**EA**	**RA**	**OR**	**95% CI**	***p*-Value**	**OR**	**95% CI**	***p*-Value**	**OR**	***p*-Value**
GALNT2	rs4846913	A	C	0.739	0.575 to 0.951	**0.019**	0.928	0.754 to 1.142	0.480	0.838	0.118
GALNT2	rs35498929	T	C	0.858	0.592 to 1.244	0.419	0.890	0.642 to 1.234	0.485	0.876	0.292
GALNT2:RP5-956O18.3	rs3213497	T	C	1.099	0.771 to 1.567	0.602	0.776	0.557 to 1.081	0.133	0.918	0.625
IFT172/NRBP1	rs67086575	G	A	0.857	0.598 to 1.226	0.398	0.934	0.703 to 1.242	0.640	0.903	0.372

Nephropathy					DiaGene			Hoorn DCS		Meta-Analysis	
Locus	**rsID**	**EA**	**RA**	**OR**	**95% CI**	***p*-Value**	**OR**	**95% CI**	***p*-Value**	**OR**	***p*-Value**
GALNT2	rs4846913	A	C	0.949	0.760 to 1.184	0.641	1.032	0.902 to 1.180	0.651	1.008	0.889
GALNT2	rs35498929	T	C	0.996	0.726 to 1.367	0.982	0.889	0.721 to 1.097	0.272	0.920	0.351
GALNT2:RP5-956O18.3	rs3213497	T	C	0.883	0.639 to 1.220	0.450	0.776	0.628 to 0.958	0.056	0.806	**0.017**
IFT172/NRBP1	rs67086575	G	A	0.868	0.639 to 1.189	0.366	1.118	0.938 to 1.333	0.212	1.018	0.884
Neuropathy					**DiaGene**			**Hoorn DCS**		**Meta-Analysis**	
GALNT2	rs4846913	A	C	1.254	0.935 to 1.681	0.130	NA	NA	NA	NA	NA
GALNT2	rs35498929	T	C	1.350	0.895 to 2.038	0.153	NA	NA	NA	NA	NA
GALNT2:RP5-956O18.3	rs3213497	T	C	1.182	0.779 to 1.795	0.432	NA	NA	NA	NA	NA
IFT172/NRBP1	rs67086575	G	A	0.978	0.656 to 1.459	0.914	NA	NA	NA	NA	NA
Macrovascular Complications					**DiaGene**			**Hoorn DCS**		**Meta-Analysis**	
GALNT2	rs4846913	A	C	1.091	0.911 to 1.307	0.343	1.036	0.781 to 1.374	0.806	1.075	0.354
GALNT2	rs35498929	T	C	0.915	0.708 to 1.182	0.495	1.255	0.844 to 1.866	0.262	1.031	0.840
GALNT2:RP5-956O18.3	rs3213497	T	C	0.990	0.760 to 1.289	0.939	1.216	0.831 to 1.780	0.314	1.059	0.606
IFT172/NRBP1	rs67086575	G	A	0.947	0.743 to 1.208	0.663	0.867	0.586 to 1.284	0.476	0.924	0.455

Beta represents the change in prevalence of the respective complication per increase of the effect allele. Bold for $p < 0.05$. EA, effect allele; RA, reference allele; NA, not available.

Table S5. Associations of lead SNPs with prevalent complications of type 2 diabetes, sensitivity analyses.

Retinopathy

				Sens. Lipid Meds								Sens. Insulin use							
				DiaGene				***Hoorn DCS***				***DiaGene***				***Hoorn DCS***			
				95% C.I. for OR				95% C.I. for OR				95% C.I. for OR				95% C.I. for OR			
uniqID	rsID (dbSNP build 146)	EA	RA	OR	lower	upper	p-value	OR	lower	upper	p-value	OR	lower	upper	p-value	OR	lower	upper	p-value
1:230294715:A:C	rs4846913	A	C	0.722	0.559	0.932	**0.012**	0.927	0.753	1.142	0.475	0.865	0.685	1.092	0.223	0.937	0.763	1.150	0.532
1:230286016:C:T	rs35498929	T	C	0.838	0.576	1.219	0.356	0.886	0.639	1.229	0.468	0.859	0.607	1.216	0.393	0.909	0.658	1.256	0.563
1:230416320:C:T	rs3213497	T	C	1.118	0.782	1.599	0.540	0.781	0.561	1.088	0.144	1.077	0.765	1.517	0.670	0.765	0.552	1.061	0.109
2:27686480:A:G	rs67086575	G	A	0.845	0.590	1.211	0.360	0.946	0.711	1.258	0.703	0.851	0.607	1.193	0.350	0.933	0.703	1.240	0.634

Nephropathy

				Sens. Lipid Meds								Sens. Insulin use							
				DiaGene				***Hoorn DCS***				***DiaGene***				***Hoorn DCS***			
				95% C.I. for OR				95% C.I. for OR				95% C.I. for OR				95% C.I. for OR			
uniqID	rsID (dbSNP build 146)	EA	RA	OR	lower	upper	p-value	OR	lower	upper	p-value	OR	lower	upper	p-value	OR	lower	upper	p-value
1:230294715:A:C	rs4846913	A	C	0.947	0.759	1.183	0.633	1.032	0.902	1.181	0.645	0.988	0.794	1.230	0.916	1.034	0.905	1.181	0.626
1:230286016:C:T	rs35498929	T	C	0.995	0.725	1.365	0.974	0.889	0.721	1.097	0.274	0.991	0.726	1.353	0.956	0.907	0.737	1.115	0.355
1:230416320:C:T	rs3213497	T	C	0.884	0.639	1.222	0.455	0.774	0.626	0.956	**0.017**	0.874	0.633	1.207	0.413	0.817	0.666	1.003	0.054
2:27686480:A:G	rs67086575	G	A	0.867	0.638	1.179	0.363	1.114	0.935	1.328	0.229	0.861	0.636	1.165	0.331	1.127	0.947	1.340	0.177

Neuropathy

Sens. Lipid Meds

Sens. Insulin use

				DiaGene				***Hoorn DCS***				***DiaGene***				***Hoorn DCS***			
					95% C.I. for OR				95% C.I. for OR				95% C.I. for OR				95% C.I. for OR		
uniqID	rsID (dbSNP build 146)	EA	RA	OR	lower	upper	p-value	OR	lower	upper	p-value	OR	lower	upper	p-value	OR	lower	upper	p-value
1:230294715:A:C	rs4846913	A	C	1.257	0.936	1.687	0.128	NA	NA	NA	NA	1.253	0.935	1.681	0.132	NA	NA	NA	NA
1:230286016:C:T	rs35498929	T	C	1.353	0.895	2.043	0.151	NA	NA	NA	NA	1.316	0.870	1.992	0.194	NA	NA	NA	NA
1:230416320:C:T	rs3213497	T	C	1.182	0.779	1.796	0.432	NA	NA	NA	NA	1.181	0.775	1.799	0.439	NA	NA	NA	NA
2:27686480:A:G	rs67086575	G	A	0.978	0.656	1.460	0.914	NA	NA	NA	NA	1.022	0.683	1.528	0.917	NA	NA	NA	NA

Macrovascular complications

Sens. Lipid Meds

Sens. Insulin use

				DiaGene				***Hoorn DCS***				***DiaGene***				***Hoorn DCS***			
					95% C.I. for OR				95% C.I. for OR				95% C.I. for OR				95% C.I. for OR		
uniqID	rsID (dbSNP build 146)	EA	RA	OR	lower	upper	p-value	OR	lower	upper	p-value	OR	lower	upper	p-value	OR	lower	upper	p-value
1:230294715:A:C	rs4846913	A	C	1.120	0.932	1.345	0.226	1.045	0.786	1.389	0.761	1.102	0.920	1.319	0.292	1.061	0.806	1.396	0.672
1:230286016:C:T	rs35498929	T	C	0.952	0.734	1.234	0.710	1.263	0.844	1.889	0.256	0.911	0.705	1.178	0.476	1.289	0.872	1.904	0.202
1:230416320:C:T	rs3213497	T	C	0.951	0.727	1.243	0.711	1.194	0.810	1.761	0.371	0.994	0.763	1.295	0.962	1.231	0.852	1.780	0.268
2:27686480:A:G	rs67086575	G	A	0.986	0.770	1.262	0.911	0.828	0.556	1.231	0.351	0.945	0.741	1.205	0.648	0.860	0.580	1.274	0.452

Bold for p<0.05, bold and underlined for p<0.05/4=0.013; EA, effect allele; RA, reference allele

The variant rs3213497-T located at an exonic non-coding RNA region in the GALNT2:RP5-956O18.3 gene, previously associated with apo-$CIII_{0a}$[221], was associated at a nominal significance level with a decreased incidence of retinopathy in our meta-analysis of the DiaGene and Hoorn DCS in Models 1 and 2 (HR = 0.830, p = 0.009 and HR = 0.834, p = 0.012 resp.) (**Table 6 and Table 7**). This association remained after the sensitivity analysis for lipid-lowering medication and insulin use, although without Bonferroni significance level (**Supplementary Table S6**). The same variant rs3213497-T was nominally significantly associated with decreased prevalence of nephropathy in the meta-analysis with OR = 0.806, p = 0.017 (**Table 5**).

Table 6. Associations of genetic variants with incident complications of type 2 diabetes, Model 1 (adjusted for age and sex).

Retinopathy					DiaGene			Hoorn DCS		Meta-Analysis	
Locus	**rsID**	**EA**	**RA**	**HR**	**95% CI**	***p*-Value**	**HR**	**95% CI**	***p*-Value**	**HR**	***p*-Value**
GALNT2	rs4846913	A	C	0.992	0.778 to 1.266	0.949	1.100	0.997 to 1.214	0.057	1.084	0.082
GALNT2	rs35498929	T	C	0.724	0.494 to 1.063	0.099	1.106	0.960 to 1.275	0.162	0.931	0.732
GALNT2:RP5-956O18.3	rs3213497	T	C	0.696	0.468 to 1.037	0.075	0.851	0.734 to 0.988	**0.034**	0.830	**0.009**
IFT172/NRBP1	rs67086575	G	A	0.916	0.654 to 1.281	0.607	1.026	0.902 to 1.168	0.693	1.011	0.855
Nephropathy					**DiaGene**			**Hoorn DCS**		**Meta-Analysis**	
GALNT2	rs4846913	A	C	0.862	0.696 to 1.067	0.173	0.989	0.906 to 1.081	0.814	0.957	0.453
GALNT2	rs35498929	T	C	1.049	0.774 to 1.423	0.756	0.952	0.831 to 1.091	0.480	0.968	0.601
GALNT2:RP5-956O18.3	rs3213497	T	C	1.069	0.796 to 1.436	0.657	0.931	0.817 to 1.061	0.284	0.952	0.422
IFT172/NRBP1	rs67086575	G	A	0.961	0.719 to 1.286	0.791	0.987	0.876 to 1.112	0.828	0.983	0.767
Neuropathy					**DiaGene**			**Hoorn DCS**		**Meta-Analysis**	
GALNT2	rs4846913	A	C	1.040	0.827 to 1.308	0.735	NA	NA	NA	NA	NA
GALNT2	rs35498929	T	C	1.258	0.897 to 1.766	0.184	NA	NA	NA	NA	NA
GALNT2:RP5-956O18.3	rs3213497	T	C	0.974	0.683 to 1.388	0.883	NA	NA	NA	NA	NA
IFT172/NRBP1	rs67086575	G	A	0.857	0.617 to 1.191	0.358	NA	NA	NA	NA	NA
Macrovascular Complications					**DiaGene**			**Hoorn DCS**		**Meta-Analysis**	

GALNT2	rs4846913	A	C	1.020	0.731 to 1.424	0.906	1.020	0.892 to 1.166	0.773	1.020	0.751
GALNT2	rs35498929	T	C	0.954	0.589 to 1.548	0.850	1.128	0.930 to 1.370	0.222	1.103	0.287
GALNT2:RP5-956O18.3	rs3213497	T	C	1.313	0.823 to 2.096	0.253	0.885	0.722 to 1.086	0.242	1.018	0.926
IFT172/NRBP1	rs67086575	G	A	0.608	0.355 to 1.040	0.069	0.985	0.822 to 1.180	0.867	0.830	0.418

Beta represents the change of incidence of the respective complication per increase of the effect allele. Bold for $p < 0.05$, bold and underlined for $p < 0.013$. EA, effect allele; RA, reference allele; NA, not available.

Table 7. Associations of genetic variants with incident complications of type 2 diabetes, Model 2 (adjusted for age, sex, duration of diabetes, and HbA1c).

Retinopathy					**DiaGene**			**Hoorn DCS**		**Meta-Analysis**	
Locus	**rsID**	**EA**	**RA**	**HR**	**95% CI**	***p*-Value**	**HR**	**95% CI**	***p*-Value**	**HR**	***p*-Value**
GALNT2	rs4846913	A	C	0.884	0.684 to 1.142	0.344	1.081	0.978 to 1.194	0.128	1.013	0.892
GALNT2	rs35498929	T	C	0.773	0.522 to 1.145	0.199	1.110	0.963 to 1.280	0.151	0.972	0.870
GALNT2:RP5-956O18.3	rs3213497	T	C	0.741	0.496 to 1.108	0.145	0.848	0.730 to 0.985	**0.031**	0.834	**0.012**
IFT172/NRBP1	rs67086575	G	A	0.852	0.612 to 1.187	0.344	1.033	0.907 to 1.177	0.624	0.999	0.991
Nephropathy					**DiaGene**			**Hoorn DCS**		**Meta-Analysis**	
GALNT2	rs4846913	A	C	0.868	0.701 to 1.076	0.197	0.980	0.896 to 1.072	0.665	0.961	0.373
GALNT2	rs35498929	T	C	1.061	0.783 to 1.437	0.704	0.967	0.844 to 1.108	0.625	0.982	0.768
GALNT2:RP5-956O18.3	rs3213497	T	C	1.082	0.805 to 1.454	0.601	0.934	0.818 to 1.066	0.310	0.957	0.477
IFT172/NRBP1	rs67086575	G	A	0.968	0.725 to 1.293	0.827	0.981	0.869 to 1.108	0.760	0.979	0.714

Neuropathy					DiaGene			Hoorn DCS		Meta-Analysis	
Locus	**rsID**	**EA**	**RA**	**HR**	**95% CI**	***p*-Value**	**HR**	**95% CI**	***p*-Value**	**HR**	***p*-Value**
GALNT2	rs4846913	A	C	1.034	0.822 to 1.300	0.778	NA	NA	NA	NA	NA
GALNT2	rs35498929	T	C	1.285	0.917 to 1.801	0.145	NA	NA	NA	NA	NA
GALNT2:RP5-956O18.3	rs3213497	T	C	0.990	0.694 to 1.411	0.956	NA	NA	NA	NA	NA
IFT172/NRBP1	rs67086575	G	A	0.860	0.618 to 1.197	0.370	NA	NA	NA	NA	NA
Macrovascular Complications					**DiaGene**			**Hoorn DCS**		**Meta-Analysis**	
GALNT2	rs4846913	A	C	0.938	0.665 to 1.323	0.716	0.999	0.872 to 1.143	0.983	0.991	0.883
GALNT2	rs35498929	T	C	0.906	0.551 to 1.490	0.698	1.151	0.947 to 1.399	0.156	1.116	0.236
GALNT2:RP5-956O18.3	rs3213497	T	C	1.387	0.865 to 2.226	0.175	0.870	0.707 to 1.071	0.189	1.045	0.848
IFT172/NRBP1	rs67086575	G	A	0.687	0.400 to 1.179	0.173	1.002	0.836 to 1.202	0.979	0.907	0.558

Beta represents the change of incidence of the respective complication per increase of the effect allele. Bold for $p < 0.05$, bold and underlined for $p < 0.013$. EA, effect allele; RA, reference allele; NA, not available.

Table S6. Associations of lead SNPs with incident micro- and macrovascular complications of type 2 diabetes, sensitivity analyses.

Table S6. Associations of lead SNPs with incident micro- and macrovascular complications of type 2 diabetes, sensitivity analyses.

				Sens. Lipid Meds								Sens. Insulin use							
Retinopathy				***Dia***				***Hoorn***				***Dia***				***Hoorn***			
				Gene				***DCS***				***Gene***				***DCS***			
				95% C.I. for HR				95% C.I. for HR				95% C.I. for HR				95% C.I. for HR			
uniqID	rsID (dbSNP build 146)	EA	RA	HR	lower	upper	p-value	HR	lower	upper	p-value	HR	lower	upper	p-value	HR	lower	upper	p-value
1:230294715:A:C	rs4846913	A	C	0.896	0.692	1.159	0.402	1.082	0.979	1.196	0.122	0.959	0.750	1.228	0.742	1.105	1.001	1.219	**0.048**
1:230286016:C:T	rs35498929	T	C	0.783	0.529	1.160	0.222	1.104	0.957	1.273	0.174	0.714	0.485	1.053	0.089	1.114	0.966	1.284	0.138
1:230416320:C:T	rs3213497	T	C	0.739	0.494	1.104	0.140	0.850	0.732	0.988	**0.034**	0.750	0.502	1.120	0.160	0.839	0.723	0.974	**0.021**
2:27686480:A:G	rs67086575	G	A	0.866	0.620	1.209	0.398	1.048	0.920	1.194	0.479	0.925	0.656	1.306	0.658	1.023	0.898	1.165	0.735

				Sens. Lipid Meds								Sens. Insulin use							
Nephropathy				***Dia***				***Hoorn***				***Dia***				***Hoorn***			
				Gene				***DCS***				***Gene***				***DCS***			
				95% C.I. for HR				95% C.I. for HR				95% C.I. for HR				95% C.I. for HR			
uniqID	rsID	EA	RA	HR	lower	upper	p-value	HR	lower	upper	p-value	HR	lower	upper	p-value	HR	lower	upper	p-value
1:230294715:A:C	rs4846913	A	C	0.877	0.707	1.088	0.234	0.981	0.897	1.074	0.683	0.862	0.696	1.067	0.173	0.990	0.906	1.082	0.828
1:230286016:C:T	rs35498929	T	C	1.081	0.798	1.464	0.617	0.968	0.844	1.109	0.637	1.046	0.771	1.419	0.772	0.954	0.832	1.092	0.492
1:230416320:C:T	rs3213497	T	C	1.067	0.794	1.433	0.669	0.930	0.815	1.062	0.283	1.072	0.798	1.440	0.646	0.931	0.816	1.061	0.282
2:27686480:A:G	rs67086575	G	A	0.974	0.728	1.303	0.858	0.977	0.865	1.103	0.704	0.962	0.718	1.288	0.792	0.986	0.875	1.112	0.823

Neuropathy

				Sens. Lipid Meds								Sens. Insulin use							
				Dia				***Hoorn***				***Dia***				***Hoorn***			
				Gene				***DCS***				***Gene***				***DCS***			
				95% C.I. for HR				95% C.I. for HR				95% C.I. for HR				95% C.I. for HR			
uniqID	rsID	EA	RA	HR	lower	upper	p-value	HR	lower	upper	p-value	HR	lower	upper	p-value	HR	lower	upper	p-value
1:230294715:A:C	rs4846913	A	C	1.005	0.794	1.272	0.968	NA	NA	NA	NA	1.046	0.831	1.316	0.702	NA	NA	NA	NA
1:230286016:C:T	rs35498929	T	C	1.273	0.906	1.787	0.164	NA	NA	NA	NA	1.257	0.896	1.763	0.185	NA	NA	NA	NA
1:230416320:C:T	rs3213497	T	C	0.977	0.684	1.396	0.897	NA	NA	NA	NA	0.970	0.681	1.383	0.868	NA	NA	NA	NA
2:27686480:A:G	rs67086575	G	A	0.854	0.613	1.189	0.349	NA	NA	NA	NA	0.850	0.611	1.183	0.336	NA	NA	NA	NA

Cardiovascular disease

				Sens. Lipid Meds								Sens. Insulin use							
				Dia				***Hoorn***				***Dia***				***Hoorn***			
				Gene				***DCS***				***Gene***				***DCS***			
				95% C.I. for HR				95% C.I. for HR				95% C.I. for HR				95% C.I. for HR			
uniqID	rsID	EA	RA	HR	lower	upper	p-value	HR	lower	upper	p-value	HR	lower	upper	p-value	HR	lower	upper	p-value
1:230294715:A:C	rs4846913	A	C	0.942	0.668	1.329	0.735	0.998	0.872	1.143	0.982	1.005	0.718	1.408	0.975	1.020	0.892	1.166	0.773
1:230286016:C:T	rs35498929	T	C	0.911	0.554	1.498	0.712	1.151	0.947	1.398	0.157	0.934	0.575	1.517	0.782	1.131	0.932	1.374	0.213
1:230416320:C:T	rs3213497	T	C	1.382	0.861	2.219	0.180	0.870	0.707	1.071	0.190	1.297	0.814	2.066	0.275	0.882	0.719	1.082	0.229
2:27686480:A:G	rs67086575	G	A	0.691	0.402	1.187	0.180	1.003	0.836	1.204	0.972	0.628	0.365	1.080	0.093	0.985	0.822	1.181	0.870

Bold for p<0.05, bold and underlined for p<0.05/4=0.013, EA, effect allele; RA, reference allele

DISCUSSION

In the present study, we found that apo-CIII glycosylation (specifically, sialylation) and the linked *GALNT2*-gene variant were associated with the prevalence and incidence of diabetic retinopathy. This suggests that glycosylation determines apo-CIII function and decreased glycosylation of apo-C-III contributes to the development of retinopathy. Further, apo-CIII glycosylation was associated with prevalent diabetic neuropathy and macrovascular complications.

Type 2 diabetes is characterised by insulin resistance and an inadequate compensatory insulin secretory response, resulting in chronic hyperglycaemia[219]. Glucose and insulin oppositely modulate the expression of apo-CIII[125, 224]. This could explain the high apo-CIII secretion rate in the presence of type 2 diabetes[225]. So far, no other studies have investigated the relationships between apo-CIII glycosylation, its genetic background, and complications of type 2 diabetes. Here, the current study findings are discussed for each complication and potential underlying mechanisms will be hypothesised considering the limited pathophysiological information that is available in the current literature.

The most pronounced finding of this study was the connection of *GALNT2*, apo-CIII glycosylation and diabetic retinopathy (**Figure 2**). Previously, we identified a genetic variant rs4846913-A, located in the *GALNT2* gene, that was negatively associated with non-glycosylated apo-CIII (apo-$CIII_{0a}$)[221]. The A-allele of this rs4846913 variant is known to increase the expression of the *GALNT2*-gene[194], which encodes a GalNAc-transferase that initiates mucin-type *O*-glycosylation of peptides, such as apo-CIII[175]. The rs4846913-A variant comes with higher glycosylation levels of apo-CIII. Although only within the DiaGene cohort, this variant was negatively associated with retinopathy. In line, increased apo-$CIII_{0a}$ was associated with the prevalence of retinopathy. Replication in a larger cohort is needed to confirm whether increased activity of *GALNT2* reduces the risk of retinopathy through glycosylation of apo-CIII. This might be an interesting target for prevention or treatment of diabetic retinopathy.

Diabetes complications develop through an interplay of risk factors, including glucose and lipid metabolism, in which apo-CIII plays a role. Previously, we have analysed apo-CIII glycosylation profiles in over 700 people without diabetes[179]. Our findings supported other studies demonstrating a positive association between the apo-$CIII_1$/apo-$CIII_2$ ratio and triglyceride levels, implying the involvement of apo-CIII sialylation in an impaired triglyceride clearance[57-59]. In line, the removal of sialic acids from apo-CIII by neuraminidase treatment decreases its potential to inhibit LPL[135]. A recent study by Kegulian et al. has shown that the two sialylated glycoforms (apo-$CIII_1$ and apo-$CIII_2$) are cleared differently by hepatic receptors, specifically heparan sulphate proteoglycans (HSPGs) such as syndecan-1 (SDC1), LDL-receptor (LDLR), and LDL receptor-related protein 1 (LRP1)[59]. In people with type 2 diabetes, the LDL particles carry more apo-$CIII_2$ and less apo-$CIII_1$ than in those without diabetes[136]. Molecular mechanisms behind the association of apo-$CIII_1$ and apo-$CIII_2$ with triglyceride levels and vascular disease have not been fully elucidated. Mauger et al. found a positive association of small dense LDL with the production rate of apo-$CIII_2$[161]. Hiukka et al. reported an essential role of sialylation for the proinflammatory effect of LDL-bound apo-CIII on human aortic endothelial cells (HEACs). They found an increased immune response of HAECs after incubation with LDL containing apo-$CIII_2$, while the immune response after incubation with LDL containing apo-$CIII_0$ or apo-$CIII_1$ did not differ from apo-CIII-free LDL[136]. It is difficult to explain how the effects on large vessels translate to small vessel disease. Nevertheless, inflammation and dyslipidaemia are known risk factors for micro- and macrovascular complications of diabetes[226]. Moreover, higher serum sialic acid levels have been associated with inflammation[227] and coronary artery disease[228]. An increased immune response-related change in sialylation might explain our positive association of apo-$CIII_2$ with retinopathy. Taken together, the association of apo-CIII sialylation with retinopathy might reflect causality. Nevertheless, the current literature does not provide a complete explanation of the pathophysiological pathway of this association.

A negative association of rs3213497-T and apo-$CIII_1$ with nephropathy was found. Notably, rs3213497-T was associated with decreased apo-$CIII_{0a}$ levels at a suggestive significance level ($p = 1.30 \times 10^{-7}$), but not with apo-$CIII_1$ in our previous GWAS[221]. Here, only apo-$CIII_1$, and not apo-$CIII_{0a}$, presented a suggestive significance level negative association with nephropathy. The negative association of apo-$CIII_1$ with nephropathy was in line with an increase in the apo-$CIII_2$/Apo-$CIII_1$ ratio in VLDL reported in a small sample of chronic kidney disease[229]. The associations of apo-CIII glycosylation with nephropathy did not reach significance after Bonferroni correction. Nevertheless, the directions of effect were similar to the associations of apo-CIII glycosylation with retinopathy, which is also a micro-vascular complication of diabetes, with overlapping risk factors[226].

Studied only in a hospital-treated subgroup of the DiaGene study, monosialylated apo-CIII (apo-$CIII_1$) levels were positively associated with neuropathy. This trend remained in all subsequent models, independent of the duration of diabetes, glucose metabolism, the use of statins or fibrates, or insulin. Neuropathy showed an opposite direction of effect compared to retinopathy and nephropathy. Possibly, the association with apo-$CIII_1$ was driven by differences in disease severity or other comorbidities of this hospital-treated subgroup.

Non-glycosylated apo-CIII (apo-$CIII_{0a}$) was negatively associated with the prevalence of macrovascular complications. This association lost significance after the addition of lipid-reducing agents to the statistical model. The use of lipid-reducing agents was associated with lower relative levels of apo-$CIII_{0a}$ (**Supplementary Table S4**). Possibly, the use of lipid-lowering medication alters the proportion of apo-$CIII_{0a}$ and drives the association of apo-$CIII_{0a}$ with macrovascular disease.

As for the strengths of this study, this is the first report investigating associations of apo-CIII glycosylation with micro- and macrovascular complications in type 2 diabetes. We used a high-resolution MS method to assess the apo-CIII glycosylation patterns in blood plasma within a large cohort of type 2 diabetes patients from all lines of care and

replicated our genetic findings in a second cohort. The DiaGene cohort served as a high-risk population for discovery analyses, while genetic associations could even be replicated in the relatively lower-risk population of Hoorn DCS. Many clinical features of these patients were available for the analysis, allowing us to correct for possible confounders. Furthermore, we were able to investigate findings on glycosylation in the light of genetic background derived from GWASs, allowing us to assess a potential direction of effect and causality for some of the associations.

Some limitations of our study remain. The most important limitation is the amount of prospective data: the sample size for the prospective analysis was small and we had a follow-up with a median of 7.8 and 9.0 years in the DiaGene and Hoorn DCS study respectively. Due to the relatively small effect sizes, strong conclusions on effect sizes cannot be drawn. Also, the data on neuropathy were incomplete, resulting in low power for the analyses of neuropathy. The apo-CIII glycosylation profiling was performed with blood samples collected at baseline, so we were not able to assess glycosylation changes over time. Furthermore, absolute apo-CIII plasma levels were not measured; therefore, adjustment for these levels was not possible. Nevertheless, expression rates of apo-$CIII_1$ and apo-$CIII_2$ in humans are comparable[161], and the distribution of glycoforms in plasma is stable in variable apo-CIII concentrations in young, healthy men[230]. In coronary artery disease patients, apo-$CIII_2$ remains stable across apo-CIII plasma concentrations, while apo-$CIII_{0c}$ decreased and apo-$CIII_1$ increased with increasing total apo-CIII[150]. Adjustment for total plasma apo-CIII concentrations did not affect the association of apo-$CIII_2$ with plasma triglycerides, according to Koska et al.[57]. Insulin levels were not available which would have been interesting in light of the effects of apo-CIII on beta cell function. Finally, the cohort was mainly of European descent; therefore, we cannot generalise our findings to other ethnic groups.

In conclusion, our findings indicate a relationship between apo-CIII glycosylation and retinopathy, neuropathy, and macrovascular complications. In addition, a genetic

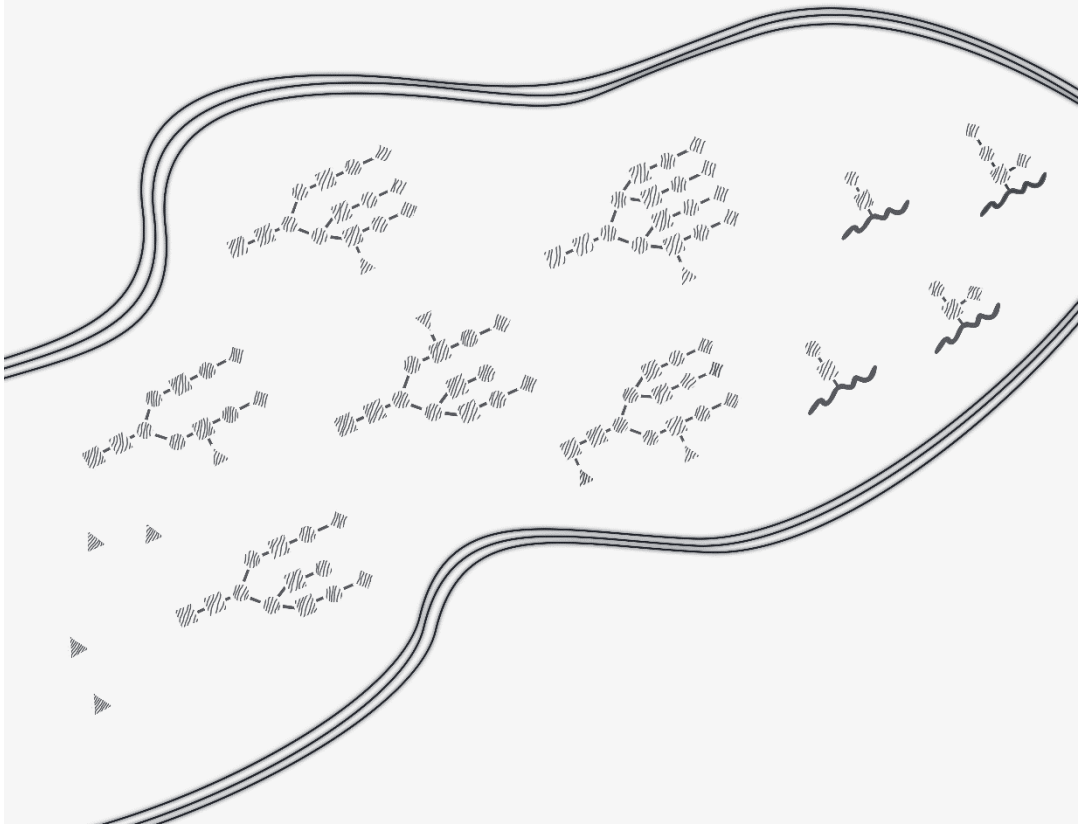
variant in the *GALNT2*-gene, previously linked to increased glycosylation of apo-CIII, was found to be negatively associated with retinopathy. Future research to further investigate the possible causal pathway of retinopathy development through aberrant apo-CIII glycosylation is warranted. Together, the current study findings suggest that apo-CIII glycosylation should be considered as a potential diagnostic and therapeutic target for diabetes complications.

Acknowledgements

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 722095 "GlySign".

The authors would like to thank the participants and staff of Diabetes Care System West-Friesland and the DiaGene Study for their cooperation and support.

CHAPTER 7



7. Discussion and perspectives

Research presented in this thesis was performed as a part of the GlySign consortium, a highly collaborative programme between academia and industry initiated in 2016, which aimed at the advancement of medical glycomics and the establishment of novel biomarkers and diagnostic tools for precision medicine, within the therapeutic areas of allo- and autoimmune diseases, prostate cancer and diabetes. The initiation of this research was motivated by the already available research stressing the importance of glycosylation involvement in the pathophysiology of many diseases, the need of bringing robust assays for glycomics analysis in a clinical laboratory setting, and with facilitating the translation of current and new glycan-based biomarkers into the clinic.

In this thesis, development and optimization of analytical methods for the evaluation of diagnostic and prognostic glycan biomarkers were described. A range of challenges are faced while performing such evaluation, which derives from the technical limitations of analytical methods, the nature of research samples and biomolecules analysed, and with data processing. Some of these challenges were faced during the research presented in this thesis for which proposing and implementing solutions was required to secure the validity and quality of the results.

Technical obstacles occur often during research performed in academia due to the novelty of research topics. Novel analytical approaches and research discoveries often emerge from academia and are eventually adopted by relevant industrial sectors. A demand for new therapies, including biomolecules with increasing functional and structural complexity, is high. This in turn brings about the need for innovative, higher throughput test methods, which are used to characterise these biomolecules. However, implementation of novel analytical approaches in biopharma, in particular their manufacturing sectors, is challenging. Normally, these sectors adhere to test methods proposed by regulatory bodies, which are outlined in regulatory guideline documents. In highly regulated environments, for instance GMP regulated laboratories of biologic manufacturing organisations, such test methods tend to employ well-

established and validated gold-standard analytical approaches, leaving little room for change or innovation[231-233]. Considering the increasing complexity of therapeutic molecules, this leads to a range of bottlenecks and hurdles in characterisation of biopharmaceuticals during manufacturing phases. Those aspects will be further discussed in the context of research and analytical approaches described in previous chapters of this thesis.

Glycosylation in healthcare and precision medicine

Precision medicine is an evolving field in healthcare, which is seeking to create and provide healthcare with improved diagnostic tools and therapies. New preventive approaches and highly efficient targeted treatment strategies, which are applicable to groups of individuals who meet certain characteristics, are expected to serve healthcare providers and support existing models of care. Improvement patient life quality and reducing the burden on public healthcare sectors are the major drivers of change, which is expected to occur with the introduction of precision medicine programs.

Research presented in this thesis was built on the basis of the latest developments in precision medicine, diagnostic biomarker discovery and bioanalytical innovations in the fields of glycomics and glycoproteomics [51, 82, 137]. Analytical technologies are evolving with multi-dimensional characterisation of large clinical sample cohorts becoming possible[234-236]. With developments of omics (e.g. proteomic, metabolomic, (epi-)genomic, glycomic) databases, powerful analytical methods for characterising patients, and a growing variety of software tools for processing large sets of data and data modelling, this field of science will be able to deliver new insights into novel personalised therapeutic approaches for disease prevention and treatment in healthcare. Currently, a growing number of such approaches, which employ novel types of biomarkers[237-239], are being proposed and investigated as potentially applicable in clinical practice. The majority of new diagnostic and therapeutic approaches is being proposed by research performed in academia, however, their translation into clinically relevant tools has been slow. There are challenges and bottlenecks that contribute to the slow pace and lack of success in bringing new precision medicine tools into clinical practice, which will be discussed further[240].

In **chapter two**, an inter-laboratory evaluation of an *N*-glycan biomarker for stratification of patients with HNF1A-MODY was described[52]. In this study an LC-based method was applied to measure antennary fucosylation levels in 320 cases with HNF1A-

MODY and type 2 diabetes. A major strength of this research was fit-for-purpose development of an LC-based method, which specifically targets antennary fucosylated *N*-glycans facilitated by semi-automated sample preparation and data processing. Enhanced analytical capabilities of the LC-based method allowed for an in-depth evaluation of the *N*-glycan biomarker for HNF1A-MODY (covering six single glycan traits and a derived antennary fucosylation trait), which complements discoveries from the previous studies researching this biomarker[51, 53]. A significant aspect of this research was its inter-laboratory character. A set of samples measured previously in another laboratory and tested as a part of this study (**chapter two**) was evaluated in the context of consistency of antennary fucosylation measurements and provided excellent results, as demonstrated by the correlation analysis for three best performing traits. The inter-laboratory character of this study brings an important value to the field of precision medicine and glycomics. Glycomics is a relatively young field and this study is the first of its kind in regards to inter-laboratory evaluation of a single *N*-glycan stratification biomarker. Lack of such inter-laboratory comparisons and the consequent lack of evidence for the in-depth performance of novel biomarkers results in little interest in commercialisation of such biomarker-based tests thus their poor translation into clinical practice.

Unfortunately, the use of LC-based methods for clinical testing is poorly recognised in clinical practice. Therefore, the chances of applying such methods for clinical testing are low. In response to this situation, an enzymatic plate-based assay was developed, which is described in **chapter three**[55]. This assay was developed on the basis of research results on *N*-glycan biomarker for HNF1A-MODY using LC methods[51-53]. A significant consideration in designing this plate-based assay was to remove major bottlenecks, such as expensive instrumentation, low throughput and data processing, which limit translation of LC methods into clinically-relevant tests. The diagnostic performance of the assay, which has been optimised, automated and validated using a thousand clinical samples, was close to the performance of LC-based assay for this biomarker with an AUC of 0.87 and an AUC of 0.90, respectively. The study described in **chapter three** is

the continuation of the evaluation of this glycan-based stratification biomarker for HNF1A-MODY and, together with the study results from **chapter two**[52], provides valuable insights into a process of glycan biomarker evaluation and validation, and the inter-laboratory and translational aspects of these studies. There is a strong evidence on applicability of the glycan biomarker-based diagnostic assay for selected patients suspected of having MODY, which was presented in **chapter two and three**. However, a low frequency of HNF1A-MODY in the general population (approximately 1–4% of diabetes cases)[241], rather low risk of misdiagnosis-related consequences (commonly misdiagnosed as type 2 diabetes)[81] and the on-going crisis in public healthcare sectors diminish the chances of bringing this assay into clinics. Nevertheless, this study can serve as a proposition of a workflow for evaluation of novel biomarkers and design of an assay, which is a prototype of a diagnostic/screening test with a potential to be further developed and applied in clinical practice. Such studies are expected to facilitate glycan biomarker research by opening the door for similar studies testing technical aspects of novel biomarkers, such as their performance across laboratories.

A significant aspect, which has to be considered while commercialisation of diagnostic tests and is typically not taken into consideration in academic research on clinical biomarker testing technologies, is adherence to *in vitro* diagnostic (IVD) regulations. IVD regulations are guidelines that establish the safety and effectiveness standards for diagnostic medical devices, enforced by regulatory bodies such as the FDA and EMA. While both ISO 9001:2015 and ISO 13485:2016 are international standards that provide guidelines for quality management systems (QMS), ISO 13485:2016 is more specific to the medical device industry. ISO 13485:2016 includes additional requirements for manufacturers of IVDs to ensure that medical device products meet appropriate standards of safety and performance, outlining a commitment to continuous improvement, risk management, and the documentation of processes and procedures. The ISO 13485 framework includes a set of standards for the development, production, and distribution of these products. Compliance with the IVD regulations is critical to obtain regulatory approvals. Adhering to these regulations ensures the quality and

reliability of biomarker-based tests, strengthening trust among healthcare providers and patients, and ultimately promoting adoption and usage of these tests in clinical practice. Of note, machine learning and AI-based modelling approaches that are developed with an intention to be applied for clinical decision-making and diagnosis also fall under medical devices regulations thus are controlled by the ISO 13485 framework. One of the requirements stated in the ISO 13485 framework is the implementation of Quality Assurance systems during R&D stages of such biomarker- or software-based tests. Fulfilling this requirement is critical, however, brings an overall escalation of development costs, and subsequently the costs of the final product. This aspect is considered as a significant impediment in the translation of novel biomarkers and diagnostic approaches into commercially-available clinical tests[242, 243].

Cohort case-control studies are a type of observational study, in which individuals are observed or certain outcomes are measured. In experimental studies researchers would introduce some form of intervention meaning that one or more factors are altered, and their effects examined. In observational studies, scientific questions are addressed by finding and evaluating associations, such as between either disease and exposure or analytically measurable changes that occur in relation to characterised clinical factors: a diagnosed disease or disease-related complication. Such analytically measurable changes that are found through association analysis could be further evaluated as potential new biomarkers[244].

Three well described sample cohorts were used in the studies described in **chapters two - six** of this thesis. A cohort of cases with HNF1A-MODY and type two diabetes collected in the UK and Croatia was studied in **chapter two and three[52, 55]**. In the following chapters, the DiaGene cohort was analysed to obtain apolipoprotein CIII (apoC-III) *O*-glycosylation profiles[141]. These profiles were then subjected to the association analysis, the outcomes of which were meta-analysed using GWAS data generated from the Hoorn DCS cohort (**chapter five and six)**[180, 245]. These three sample cohorts had been previously characterised in multiple other studies using various

analytical approaches to obtain GWAS and *N*-glycome profile datasets for thousands of individuals that provided blood plasma samples for these studies[28, 33, 246]. In **chapter four**, apoC-III glycosylation profiles were measured by a highly-automated method employing a liquid-handling platform and MALDI-FTICR MS[179]. A similar workflow had been previously applied to obtain *N*-glycosylation profiles from big cohorts of blood plasma samples[28, 247]. Unlike *N*-glycosylation profiles, which are stable and not affected by freeze-thaw cycles[248, 249], protein analyses reveal other challenges. Fresh human samples would be the most desirable in certain analyses, however, this is a rather unusual case scenario. Sample cohort collection is a lengthy process initiated by multi-centre collaborative efforts. Moreover, most of large sample cohorts are often used multiple times in various studies, which requires their long-term storage and transfer. It has been shown in the past that the freeze-thaw cycles cause greater protein degradation than due to long-term storage at -70^{o}C[250] and sample stability assessment has always been a part of bioanalytical method validation. Jian at al. tested the influence of long-term storage, freeze-thaw cycles and room temperature exposure on plasma samples in a study on apoC-III and it was reported that ratios of apoC-III proteoforms remain constant regardless of the exposure to various sample-handling conditions[251]. Despite the good stability of apoC-III proteoforms in stored plasma samples, these analytes might undergo certain modifications. For instance, repeated freeze-thaw cycles have an effect on protein oxidation in plasma samples. Due to the nature of the study analysing a glycoprotein and considering a large number of cases (n = 771, n = 2318 and n = 1571 in **chapter four, five and six**, respectively), new challenges arose, which put a new perspective on sample preparation optimisation and dataset processing in this study. Those challenges also drew attention to data quality of this and similar studies on apoC-III using MALDI MS.

Clinical biomarker discovery and validation relies on studying large number of samples, which require repeatable high-throughput analytical methods. The study described in **chapter four** addresses challenges posed by large-scale clinical research, overcoming a problem of plasma samples heterogeneity due to varying levels of protein oxidation

arising from storage conditions and sample processing[179]. A major strength of the method presented in this chapter is reduced methionine oxidation heterogeneity, which allowed for high-throughput analysis of apo-CIII glycosylation. The limitations of this method arise from the use of MALDI-FTICR MS which is high-end instrumentation and is not commonly found in research laboratories. Sophisticated analytical instrumentation is commonly found in centralised research facilities, which are accessible to many scientists (through collaborations) working in different research fields. These technologies, which are not meant for routine clinical applications, are used to overcome the limitations provided by commonly used, less performing instrumentation. 15T MALDI-FT-ICR MS provides mass spectra of superior quality compared to MALDI-TOF MS although the latter technology is more diffuse. In MALDI-TOF MS measurements, apoC-III glycoforms are detected as broad peaks due to low resolution and apoC-III oxidoforms are not resolved. This leads to the broadening and distortion of the apoC-III glycoform signals which can eventually overlap with other apoC-III proteoforms or other proteins affecting the quantification. On the basis of previous studies on apoC-III *O*-glycosylation by MALDI-TOF MS and the study described in **chapter four**, the severity of this problem depends on the method used for apoC-III enrichment from plasma/serum samples. The chance of overlaps between signals is higher when apoC-III is enriched using solid-phase extraction (SPE), such as C18- or C4-based SPE. In this case, other small plasma/serum proteins are co-enriched and measured in the same m/z-range of apoC-III glycoforms. The problem of overlaps is reduced, but not solved when a more specific enrichment method is used, for example using the immunocapture of apoC-III. Trenchevska and co-workers reported on an MS-immunoassay for the analysis of apoC-III proteoforms including glycoforms and truncated forms[252]. In this study, mono-oxidation of apoC-III glycoforms and overlapping of apo-$CIII_1$ with its truncated form apoC-$III_{1\text{-}A}$ are clearly visible. Since both oxidation and truncation of apoC-III glycoforms can occur to a different extent even in fresh samples, overlapping species might affect the accurate quantification of apoC-III glycoforms. Unlike the MS data curation step applied in the study described in **chapter**

four, the evaluation of peak shape of apoC-III glycoforms detected in hundreds of MALDI-TOF mass spectra has never been reported. Such evaluation would highlight the limitation of MALDI-TOF MS compared to MALDI-FT-ICR MS regarding quantification accuracy of species with similar m/z-values. The quality provided by MALDI-TOF MS seems to be sufficient to report differences between the expression of the apoC-III glycoforms in current studies (e.g. researching congenital disorders affecting *O*-glycosylation)[253, 254], however, for other applications that quality may not be sufficient as it may fail to detect more subtle differences.

Genome-wide association studies (GWAS) help to identify variations, called single nucleotide polymorphisms (SNPs), in genes statistically associated with a particular trait or disease. By studying genetic variants across many genomes, these significant associations emerge, and it becomes possible to gain insights into molecular mechanisms that are likely involved in disease development and identify novel potential drug targets[255]. In **chapter five**, apoC-III *O*-glycosylation including sialylation, and a genomic variants datasets from 2318 cases with type 2 diabetes being part of the DiaGene cohort were subjected to a genome-wide association analysis[221]. Additionally, associations between genetic variants and lipid biomarkers were meta-analysed against the Hoorn DCS cohort. In **chapter six**, in addition to the associations between apolipoprotein-CIII glycosylation levels and micro- and macrovascular complications of type 2 diabetes within DiaGene cohort, the associations of apoC-III glycosylation-linked genetic variants with the prevalence and incidence of type 2 diabetes micro- and macro-vascular complications were meta-analysed against the Hoorn DCS cohort[245]. Despite the potential of GWAS findings for uncovering novel associations between genetic variants and disease or other measurable traits, there are outstanding questions in these types of studies. In the instance when identified genetic variants are located outside coding regions, which most likely will not lead to direct changes in the structure of a protein, interpreting GWAS associations requires additional experimental validation. Commonly proposed approaches for the validation of GWAS associations

are combining GWAS with gene editing and cellular phenotyping or integrating GWAS with various omics data to investigate the functional manifestations[256].

Due to the immense complexity and inter-connectivity of molecular mechanisms in human physiology and, therefore, inability to fully mimic *in vivo* pathophysiological conditions under *in vitro* conditions, no functional experimental validation was performed in the studies presented in **chapter five and six[221, 245].** Nevertheless, various reports on molecular and genetic aspects of apo-CIII glycosylation, diabetes and lipid metabolism-related mechanisms aid in understanding the causality of research findings presented in these chapters. Research described in **chapter five** revealed genetic variants in *GALNT2* and *IFT172* genes, which had previously been directly or indirectly linked to *O*-glycosylation mechanisms, to be associated with apoC-III glycosylation and lipid levels[221]. It was concluded that the occurrence of these genetic variants could modulate lipid levels by altering glycosylation of apo-CIII or other proteins involved in lipid metabolism. Up to date, the presence of specific apolipoproteins on triglyceride-rich lipoprotein (TRL) particles, such as hepatic VLDL, has been recognised to regulate TRL metabolism by modulating their lipidation and secretion, lipoprotein lipase (LPL) mediated TRL hydrolysis, and TRL remnant uptake by liver[123]. There is a growing evidence of the involvement of glycosylation in critical mechanisms of lipid metabolism. For instance, the glycosylation status of apoC-III, specifically sialylation, affects the binding of TRL-rich particles to biglycan, a chondroitin or dermatan sulfate proteoglycan present in the arterial wall [136]. Further, it was demonstrated that the sialylation status of apo-CIII impacts its inhibitory capacity against LPL[135] and its affinity to hepatic receptors that play a role in TRL remnant uptake[59]. LPL, an enzyme degrading circulating triglycerides, contains two N-glycosylation sites, with glycosylation at N70 being crucial for LPL activity and intracellular trafficking[257]. Altogether, the presence or absence of varying glycoforms on both apo-CIII and LPL might significantly impact lipid metabolism through the interactions between LPL, apo-CIII containing TRL particles and hepatic TRL uptake receptors. However, the involvement of other unstudied protein intermediates, their glycosylation status or epigenetic factors cannot be dismissed. The

impact of glycosylation on lipid metabolism-related pathologic conditions might be larger than currently proposed in literature. As described in **chapter five[221],** all discovered genetic variants were directly or indirectly linked to glycosylation mechanisms, such as the initiation of mucine-type *O*-linked glycosylation for the genetic variants in *GALNT2* and subcellular trafficking between the endoplasmic reticulum and Golgi apparatus for the genetic variants in *IFT172* via a crosslink to the *NRBP1* gene. **In chapter six[245]**, interconnections between the previously studied genetic variants in *GALNT2* and *IFT172*, their effect on apo-CIII glycosylation status and development of type 2 diabetes micro- and macro-vascular complications were investigated. The most pronounced interconnection was found for retinopathy and genetic variants in *GALNT2*, suggesting that modulating apo-CIII glycosylation via *GALNT2* might be an important target for further research and the development of new preventive approaches for retinopathy. Of note, epigenetic factors play a critical role in gene activity without altering the DNA sequence[258]. Behavioural and environmental changes are known to introduce genetic modifications through epigenetic changes, with DNA methylation and non-coding RNA sequences playing key roles in the regulation of gene expression[259], which also have been specifically investigated in the pathogenesis of type 2 diabetes[260]. Moreover, certain medications used in type 2 diabetes treatment have been shown to induce epigenetic changes, potentially contributing to their therapeutic effects[261, 262]. Although sensitivity analysis and confounding approaches were applied in the studies presented in **chapters five and six**, which may partially diminish the potential effect of epigenetic changes related to such factors as medications, obesity or smoking, epigenetic aspects should be taken into account in future studies and during development of new therapeutic strategies.

Taking into consideration fit-for-purpose optimisation of the analytical method used to obtain apo-CIII *O*-glycosylation profiles and rigorous, systematic data curation applied in the studies on apo-CIII glycosylation described in **chapter four, five and six**[179, 221, 245], a new approach for the analysis and validation of novel biomarkers in large sample cohorts was proposed. By applying this approach, it was possible to assess the quality

of data. Although reporting detailed data quality metrics adds a significant value to research discoveries, it is not a standard practice and depends on awareness of research centres and scientific journals including editors and publishers. As stated earlier, MALDI-FTICR MS use can be a bottleneck in translating this type of biomarker into clinics. Nevertheless, results obtained in such large-scale studies with a focus on data quality and validity can serve as a base for further research, allowing for performing inter-study or inter-laboratory evaluations and developing simpler assays, such as the enzymatic plate-based assay proposed and described in **chapter three**[55].

In a future perspective, the collection of multi-dimensional datasets backed by strong data quality metrics will enable in-depth investigation of novel associations, for instance between glycosylation and GWAS or *N*-glycosylation and *O*-glycosylation, and facilitate studies, in which datasets (e.g. *N*-glycosylation profiles) obtained by different analytical methods or/and in various research laboratories are a subject of comparison and technical evaluation. Up to date, the performance of methods for high-throughput analysis of serum and IgG *N*-glycome have been reported in several studies, in which methods accuracy, repeatability and throughput were assessed[263, 264]. Such comparative studies are of high significance, especially in relatively young scientific fields such as glycomics. **Chapter two** presents development of a novel analytical method, which was applied to obtain antennary fucosylation profiles in blood plasma from a set of patients with HNF1A-MODY and type 2 diabetes[52]. These profiles were subjected to the comparative analysis and correlated with antennary fucosylation levels measured by a different method in another independent laboratory. By performing this research, it was possible to present the validity of both analytical approaches and confirm the validity of a glycan-based stratification biomarker, which both contribute to strengthen the value of glycosylation analysis and clinical applications of glycans.

In human (cohort) research studies, *in vivo* functional studies are usually unavailable and *in vitro* functional studies cannot fully represent the true *in vivo* scenario. In

chapter five[221], by combining GWAS and glycoproteomics data, it was possible to find a link between genetic variants and their functional manifestation that confirms that glycosyltransferases might be promising diagnostic and therapeutic targets, which is often indicated in cancer research.[265, 266] Further, in **chapter six**[245], the meta-analysis of genetic variants and type 2 diabetes complications allowed to strengthen the link between the previously investigated genetic variant in the *GALNT2* gene, its functional manifestation of apo-CIII glycosylation and the incidence of retinopathy. Altogether, the large cohort and comparative studies carry not only the advantage of statistical power and the potential to provide unique research discoveries in the field, but also valuable insights into the technical performance of analytical methods for glycosylation analysis.

Glycosylation in industry and biopharma research

Despite developments and increasing availability of diagnostic tools, which are accurate and allow early diagnosis, there is an urgent need for safer and more efficient therapeutic approaches. Taking into consideration the burden on public healthcare budget brought by the COVID-19 pandemic, increasing numbers of other high-cost chronic diseases and the economic crisis, a high demand for safer and broadly accessible therapies (including biopharmaceutical products) is warranted[267, 268].

Glycosylation is considered as one of the critical quality attributes in biologic manufacturing. Developments in IgG-type monoclonal antibodies, following bispecific antibodies and bispecific derivatives have shown that by designing and adjusting glycosylation patterns, it is possible to influence the half-life, immunogenicity and pharmacokinetics of therapeutic molecules. Glycosylation patterns have to be closely monitored throughout biologic development and manufacturing. Cell line engineering and adjusting fermentation process conditions by media supplementation approaches are common practices to control molecule glycosylation patters, alongside more sophisticated and expensive upstream-downstream processing strategies that employ downstream enzymatic remodelling and affinity-based separation of protein glycovariants[269].

Alternatively to the genetic disruption of glycan biosynthesis, the use of small molecules as glycosylation enzyme inhibitors has been proposed and evaluated by Shasha Li and co-workers[270]. This proposition falls under the media supplementation approaches and is considered operationally simple as well as cost effective, however, its application requires extensive case-to-case evaluation. The small molecules, which are applied in this approach, have to possess a range of critical characteristics - their efficiency must be maintained while the stability of expression systems and expressed proteins remain not affected. Furthermore, computationally-informed strategies can potentially be employed to increase the relative abundance of targeted glycan structures thus aid the production of more homogenous biologics with respect to

glycosylation[271]. Fisher et al. have reported a computational model, which has the potential to be utilised for glyco-engineering of biologics as well as further elucidation and advancing treatments for glycosylation-related diseases[272]. Nevertheless, achieving complete homogeneity of glycoforms of biologics remains a significant challenge.

A demand for characterisation and matching of glycosylation patterns, which is key to demonstrate batch-to-batch consistency and obtain regulatory approvals, is also increasing in a field of biosimilar manufacturing and has been proven challenging. Those challenges arise due to the increasing complexity of therapeutic biomolecules and insufficient power of current analytical test methods[273].

In **chapter two**, development of an LC-MS/MS method for analysing antennary fucosylation of blood plasma protein has been described. LC-MS is a standard test method to characterise glycosylation profiles of biologics in industry, often supported by exoglycosidases. **Chapter two** presents stages of method development, highlighting technical limitations, such as differentiating core and antennary fucosylation of *N*-glycans using LC-MS/MS without exoglycosidases, and low throughput of such methods when considering sample and data processing without the use of automated liquid-handling platforms and software. The limitations of LC-MS methods for glycan profiling, which are advised by regulatory bodies and routinely used in biologic development and manufacturing to characterise products and collect data for regulatory purposes, are significant. Those methods are not universal and often cannot fully characterise more complex therapeutic molecules, without being further optimised. Exoglycosidases, such as the one employed in the development of the LC-MS/MS method in **chapter two**, have served as a useful tool for the optimization and development of LC-MS methods for glycan profiling in biopharma, ensuring that these methods are fit for purpose. In many cases this optimisation approach may still be insufficient taking into consideration the increasing complexity of therapeutic proteins and their glycan patterns, as well as limitations of chromatographic separation (e.g. coeluting glycan

structures) applied in UHPLC systems. Thus, differentiation and quantification of critical glycosylation features, e.g. bisection or fucosylation, when *N*-glycan antennary fucosylation is also considered, may pose a challenge in the biologics manufacturing. In a study by Moran and co-workers, a more complex quantification approach of combining fluorescence and ESI mass spectrometry signals has been applied to procainamide-labelled blood plasma *N*-glycans with sialic acid linkage-specific derivatization via ethyl esterification separated using C18 reversed-phase UHPLC column[20]. This study addresses the problem of co-eluting glycan structures and benefits from differentiation of α2,3- and α2,6-sialylated *N*-glycans. Although the complexity of data analysis and the subsequent increase in time required for the analysis might limit the use of this novel quantification approach in industry, this approach can serve as an additional or orthogonal method to current analytical approaches.

As previously mentioned, low throughput of sample processing and data analysis in glycoprofiling are major bottlenecks of current standard test methods in biologic manufacturing, but also methods used during biologic discovery phases. Moreover, processing and interpreting MS and MS/MS data from glycan profiling analyses require highly specialised and trained operators, and this also contributes to the final product's cost. In **chapter three**, an exoglycosidase plate-based assay for detecting α1-3,4 fucosylation in blood plasma proteins was proposed as a diagnostic tool for patients with HNF1A-MODY[55]. Development of this type of assay, which enables the analysis of glycosylation features of proteins in an automated high-throughput manner and benefits from simplified data processing, is of great industrial importance. The employment of liquid handling platforms in industry is more and more common. Moreover, enzyme discovery together with protein engineering approaches will facilitate developments of such assays[85, 274]. In **chapter three**, α1-3,4 fucosylation levels in complex mixtures of proteins were measured using the exoglycosidase plate-based assay. Regarding the characteristics of the exoglycosidase used in this study, the enzyme allowed for measuring of α1-3,4 fucosylation levels from glycans attached to proteins in blood plasma matrix. Previously, a similar assay has been developed and

applied to IgG samples to measure *N*-glycan galactosylation levels, which also allowed for quantification of sialylation levels[31]. Considering the complexity of blood plasma samples and relatively low complexity of biologic samples, which are highly purified biomolecules and their matrix can be easily adjusted, this type of a plate-based assay might eventually find a broader application in biopharma industry. Not only can the time and cost of development and manufacturing phases be reduced, aiding delivery of new therapeutics and biosimilars thus increasing their accessibility to patients. Employment of such assays measuring absolute levels of glycosylation features will add another analytical dimension to biomolecule characterisation. This may be essential for a more comprehensive evaluation of the quality of biopharmaceutical products and to ensure their safety for use in patients..

MALDI MS, which was applied in studies described in **chapter four, five and six[179, 221, 245]**, has still rather limited application in biologics manufacturing. Peptide mapping and intact mass analysis for regulatory approval purposes are performed routinely using the gold-standard LC-MS/MS method. However, those methods still suffer from major drawbacks, for instance, introducing artificial protein modifications in peptide mapping[275]. Despite the growing throughput in areas of sample preparation and processing, which is largely due to the introduction of liquid handling platforms, data processing and data analysis remain major bottlenecks in large-scale delivery of biologics. This aspect draws attention to the availability of commercially available data processing software, which aids high-throughput processing of LC (MS) and MALDI MS data for a large number of samples and provides a range of data quality metrics, similar to the software applied in studies described in **chapters two[52], four, five and six**. Developments in commercially available semi-automated data processing software are being noticeable (e.g. BioPharma Finder by Thermo Scientific or Protein Metrics by Dotmatics), however, manual intervention is still required in the majority of glycan and peptide mapping analysis. Currently, the generation and collection of large omics datasets (genomics, transcriptomics, proteomics and metabolomics) is becoming a more common practice in industry. The growing interest of industry in multi-omics

technologies and benefits associated with their use, in regards to robust drug discovery as well as process monitoring, optimization and control, makes MALDI MS a method of interest. With the accessibility of state-of-art sequencers and mass spectrometry, and developments in the fields of bioinformatics and artificial intelligence (AI), it is now becoming possible to approach diseases by probing more complex and transient disease- and treatment-related molecular changes to select new drug targets and validate them. As described in **chapter five**[221], combining large data sets, partially generated by MALDI-FTICR MS, can lead to important discoveries linking glyco(proteo)mics and genomics. In the field of glycomics, where large data sets are being delivered but still little is known about molecular mechanisms of aberrant protein glycosylation, new discoveries on interconnections between glycomics and other types of omics are considered as drivers for establishment of new research and innovations. Since the evaluation of glycan composition and the optimization of glycosylation patterns for specific therapeutic applications are crucial during biologics delivery stages, innovations brought to the field of glycomics will contribute to the progress of both academic and industrial sectors, and thus aid the delivery of new precision therapies and biotherapeutics.

Bibliography

1. Rini, J.M., et al., *Glycosyltransferases and Glycan-Processing Enzymes*, in *Essentials of Glycobiology*, A. Varki, et al., Editors. 2022, Cold Spring Harbor Laboratory Press_Copyright © 2022 The Consortium of Glycobiology Editors, La Jolla, California; published by Cold Spring Harbor Laboratory Press; doi:10.1101/glycobiology.4e.6. All rights reserved.: Cold Spring Harbor (NY). p. 67-78.
2. Varki, A., et al., *Essentials of Glycobiology*, in *Essentials of Glycobiology*, A. Varki, et al., Editors. 2022: Cold Spring Harbor (NY).
3. Chen, B., et al., *Impact of N-Linked Glycosylation on Therapeutic Proteins.* Molecules, 2022. **27**(24).
4. Clemetson, K.J., *Blood glycoproteins*. Glycoproteins II. Vol. 29. 1997.
5. Shental-Bechor, D. and Y. Levy, *Effect of glycosylation on protein folding: A close look at thermodynamic stabilization.* Proceedings of the National Academy of Sciences, 2008. **105**(24): p. 8256-8261.
6. Van Coillie J, S.M., Bentlage AEH, de Haan N, Ye Z, Geerdes DM, van Esch WJE, Hafkenscheid L, Miller RL, Narimatsu Y, Vakhrushev SY, Yang Z, Vidarsson G, Clausen H., *Role of N-Glycosylation in FcγRIIIa interaction with IgG.* Front Immunol, 2022. **13**.
7. Peracaula, R., A. Sarrats, and P.M. Rudd, *Liver proteins as sensor of human malignancies and inflammation.* Proteomics Clin Appl, 2010. **4**(4): p. 426-31.
8. McCarthy, C., et al., *The Role and Importance of Glycosylation of Acute Phase Proteins with Focus on Alpha-1 Antitrypsin in Acute and Chronic Inflammatory Conditions.* Journal of Proteome Research, 2014. **13**(7): p. 3131-3143.
9. Dominiczak, M.H. and M.J. Caslake, *Apolipoproteins: metabolic role and clinical biochemistry applications.* Ann Clin Biochem, 2011. **48**(Pt 6): p. 498-515.
10. Savinova, O.V., et al., *Reduced apolipoprotein glycosylation in patients with the metabolic syndrome.* PLoS One, 2014. **9**(8): p. e104833.
11. Kang, S.K., et al., *The hepatitis B virus X protein inhibits secretion of apolipoprotein B by enhancing the expression of N-acetylglucosaminyltransferase III.* J Biol Chem, 2004. **279**(27): p. 28106-12.
12. Harvey, S.B., et al., *O-glycoside biomarker of apolipoprotein C3: responsiveness to obesity, bariatric surgery, and therapy with metformin, to chronic or severe liver disease and to mortality in severe sepsis and graft vs host disease.* J Proteome Res, 2009. **8**(2): p. 603-12.
13. Kondo, A., et al., *Glycopeptide profiling of beta-2-glycoprotein I by mass spectrometry reveals attenuated sialylation in patients with antiphospholipid syndrome.* J Proteomics, 2009. **73**(1): p. 123-33.
14. Qin, Y., et al., *Serum glycopattern and Maackia amurensis lectin-II binding glycoproteins in autism spectrum disorder.* Sci Rep, 2017. **7**: p. 46041.
15. Subramanian, S.P. and R.L. Gundry, *The known unknowns of apolipoprotein glycosylation in health and disease.* iScience, 2022. **25**(9): p. 105031.
16. Pabst, M., et al., *Comparison of fluorescent labels for oligosaccharides and introduction of a new postlabeling purification method.* Anal Biochem, 2009. **384**(2): p. 263-73.
17. Chen, X. and G.C. Flynn, *Analysis of N-glycans from recombinant immunoglobulin G by on-line reversed-phase high-performance liquid chromatography/mass spectrometry.* Analytical Biochemistry, 2007. **370**(2): p. 147-161.
18. Pucić, M., et al., *High throughput isolation and glycosylation analysis of IgG-variability and heritability of the IgG glycome in three isolated human populations.* Mol Cell Proteomics, 2011. **10**(10): p. M111.010090.
19. Krištić, J., et al., *Glycans are a novel biomarker of chronological and biological ages.* J Gerontol A Biol Sci Med Sci, 2014. **69**(7): p. 779-89.
20. Moran, A.B., et al., *Sialic Acid Derivatization of Fluorescently Labeled N-Glycans Allows Linkage Differentiation by Reversed-Phase Liquid Chromatography-Fluorescence Detection-Mass Spectrometry.* Anal Chem, 2022. **94**(18): p. 6639-6648.

21. Zhang, T., et al., *Development of a 96-well plate sample preparation method for integrated N- and O-glycomics using porous graphitized carbon liquid chromatography-mass spectrometry.* Mol Omics, 2020. **16**(4): p. 355-363.
22. Madunić, K., et al., *Colorectal cancer cell lines show striking diversity of their O-glycome reflecting the cellular differentiation phenotype.* Cell Mol Life Sci, 2021. **78**(1): p. 337-350.
23. Banerjee, S. and S. Mazumdar, *Electrospray ionization mass spectrometry: a technique to access the information beyond the molecular weight of the analyte.* Int J Anal Chem, 2012. **2012**: p. 282574.
24. Zenobi, R. and R. Knochenmuss, *Ion formation in MALDI mass spectrometry.* Mass Spectrometry Reviews, 1998. **17**(5): p. 337-366.
25. Reiding, K.R., et al., *High-throughput profiling of protein N-glycosylation by MALDI-TOF-MS employing linkage-specific sialic acid esterification.* Anal Chem, 2014. **86**(12): p. 5784-93.
26. de Haan, N., et al., *Glycomics studies using sialic acid derivatization and mass spectrometry.* Nat Rev Chem, 2020. **4**(5): p. 229-242.
27. Clerc, F., et al., *Plasma N-Glycan Signatures Are Associated With Features of Inflammatory Bowel Diseases.* Gastroenterology, 2018. **155**(3): p. 829-843.
28. Memarian, E., et al., *Plasma protein N-glycosylation is associated with cardiovascular disease, nephropathy, and retinopathy in type 2 diabetes.* BMJ Open Diabetes Res Care, 2021. **9**(1).
29. Lippold, S., et al., *Glycoform analysis of intact erythropoietin by MALDI FT-ICR mass spectrometry.* Anal Chim Acta, 2021. **1185**: p. 339084.
30. Krishnamoorthy, L. and L.K. Mahal, *Glycomic Analysis: An Array of Technologies.* ACS Chemical Biology, 2009. **4**(9): p. 715-732.
31. Rebello, O.D., et al., *A novel glycosidase plate-based assay for the quantification of galactosylation and sialylation on human IgG.* Glycoconj J, 2020. **37**(6): p. 691-702.
32. de Haan, N., M. Wuhrer, and L.R. Ruhaak, *Mass spectrometry in clinical glycomics: The path from biomarker identification to clinical implementation.* Clin Mass Spectrom, 2020. **18**: p. 1-12.
33. Elham, M., et al., *IgG N-glycans are associated with prevalent and incident complications of type 2 diabetes.* medRxiv, 2022: p. 2022.03.15.22272417.
34. Rudd, P., et al., *Glycomics and Glycoproteomics*, in *Essentials of Glycobiology*, A. Varki, et al., Editors. 2015, Cold Spring Harbor Laboratory Press_Copyright 2015-2017 by The Consortium of Glycobiology Editors, La Jolla, California. All rights reserved.: Cold Spring Harbor (NY). p. 653-66.
35. Nicolardi, S., et al., *Identification of new apolipoprotein-CIII glycoforms with ultrahigh resolution MALDI-FTICR mass spectrometry of human sera.* J Proteome Res, 2013. **12**(5): p. 2260-8.
36. Hu, M., et al., *Chapter One - Glycan-based biomarkers for diagnosis of cancers and other diseases: Past, present, and future*, in *Progress in Molecular Biology and Translational Science*, L. Zhang, Editor. 2019, Academic Press. p. 1-24.
37. Seeling, M., C. Brückner, and F. Nimmerjahn, *Differential antibody glycosylation in autoimmunity: sweet biomarker or modulator of disease activity?* Nat Rev Rheumatol, 2017. **13**(10): p. 621-630.
38. Joenvaara, S., et al., *Quantitative N-glycoproteomics reveals altered glycosylation levels of various plasma proteins in bloodstream infected patients.* PLOS ONE, 2018. **13**(3): p. e0195006.
39. Gornik, O., et al., *Stability of N-glycan profiles in human plasma.* Glycobiology, 2009. **19**(12): p. 1547-53.
40. Novokmet, M., et al., *Changes in IgG and total plasma protein glycomes in acute systemic inflammation.* Sci Rep, 2014. **4**: p. 4347.
41. Šimurina, M., et al., *Glycosylation of Immunoglobulin G Associates With Clinical Features of Inflammatory Bowel Diseases.* Gastroenterology, 2018. **154**(5): p. 1320-1333.e10.
42. Pavić, T., et al., *N-glycosylation patterns of plasma proteins and immunoglobulin G in chronic obstructive pulmonary disease.* J Transl Med, 2018. **16**(1): p. 323.
43. Association, A.D., *2. Classification and Diagnosis of Diabetes.* Diabetes Care, 2016. **40**(Supplement_1): p. S11-S24.
44. ADA, A.D.A., *Diagnosis and classification of diabetes mellitus.* Diabetes Care, 2009. **32 Suppl 1**(Suppl 1): p. S62-7.
45. Kavvoura, F.K. and K.R. Owen, *Maturity onset diabetes of the young: clinical characteristics, diagnosis and management.* Pediatr Endocrinol Rev, 2012. **10**(2): p. 234-42.

46. Cade, W.T., *Diabetes-related microvascular and macrovascular diseases in the physical therapy setting.* Phys Ther, 2008. **88**(11): p. 1322-35.
47. Keser, T., et al., *Increased plasma N-glycome complexity is associated with higher risk of type 2 diabetes.* Diabetologia, 2017. **60**(12): p. 2352-2360.
48. Gornik, I., et al., *Hyperglycaemia in critical illness is a risk factor for later development of type II diabetes mellitus.* Acta Diabetol, 2010. **47 Suppl 1**: p. 29-33.
49. Rudman, N., O. Gornik, and G. Lauc, *Altered N-glycosylation profiles as potential biomarkers and drug targets in diabetes.* FEBS Lett, 2019. **593**(13): p. 1598-1615.
50. Lemmers, R.F.H., et al., *IgG glycan patterns are associated with type 2 diabetes in independent European populations.* Biochim Biophys Acta Gen Subj, 2017. **1861**(9): p. 2240-2249.
51. Juszczak, A., et al., *Plasma Fucosylated Glycans and C-Reactive Protein as Biomarkers of HNF1A-MODY in Young Adult-Onset Nonautoimmune Diabetes.* Diabetes Care, 2019. **42**(1): p. 17-26.
52. Demus, D., et al., *Interlaboratory evaluation of plasma N-glycan antennary fucosylation as a clinical biomarker for HNF1A-MODY using liquid chromatography methods.* Glycoconj J, 2021. **38**(3): p. 375-386.
53. Thanabalasingham, G., et al., *Mutations in HNF1A result in marked alterations of plasma glycan profile.* Diabetes, 2013. **62**(4): p. 1329-37.
54. Tijardović, M., et al., *Fucosylated AGP glycopeptides as biomarkers of HNF1A-Maturity onset diabetes of the young.* Diabetes Res Clin Pract, 2022. **185**: p. 109226.
55. Demus, D., et al., *Development of an exoglycosidase plate-based assay for detecting alpha1-3,4 fucosylation biomarker in individuals with HNF1A-MODY.* Glycobiology, 2022. **32**(3): p. 230-238.
56. Vijayaraghavan, K., *Treatment of dyslipidemia in patients with type 2 diabetes.* Lipids Health Dis, 2010. **9**: p. 144.
57. Koska, J., et al., *Disialylated apolipoprotein C-III proteoform is associated with improved lipids in prediabetes and type 2 diabetes.* J Lipid Res, 2016. **57**(5): p. 894-905.
58. Yassine, H.N., et al., *The Association of Human Apolipoprotein C-III Sialylation Proteoforms with Plasma Triglycerides.* PLoS One, 2015. **10**(12): p. e0144138.
59. Kegulian, N.C., et al., *ApoC-III Glycoforms Are Differentially Cleared by Hepatic TRL (Triglyceride-Rich Lipoprotein) Receptors.* Arterioscler Thromb Vasc Biol, 2019. **39**(10): p. 2145-2156.
60. Tomana, M., et al., *Abnormal glycosylation of serum igg from patients with chronic inflammatory diseases.* Arthritis & Rheumatism, 1988. **31**(3): p. 333-338.
61. Seeling, M., C. Brückner, and F. Nimmerjahn, *Differential antibody glycosylation in autoimmunity: sweet biomarker or modulator of disease activity?* Nature Reviews Rheumatology, 2017. **13**(10): p. 621-630.
62. Go, M.F., R.E. Schrohenloher, and M. Tomana, *Deficient Galactosylation of Serum IgG in Inflammatory Bowel Disease: Correlation with Disease Activity.* Journal of Clinical Gastroenterology, 1994. **18**(1).
63. Parekh, R.B., et al., *Association of rheumatoid arthritis and primary osteoarthritis with changes in the glycosylation pattern of total serum IgG.* Nature, 1985. **316**(6027): p. 452-457.
64. Moore, J.S., et al., *Increased levels of galactose-deficient IgG in sera of HIV-1-infected individuals.* AIDS, 2005. **19**(4).
65. Qian, Y., et al., *Quantitative Analysis of Serum IgG Galactosylation Assists Differential Diagnosis of Ovarian Cancer.* Journal of Proteome Research, 2013. **12**(9): p. 4046-4055.
66. Holst, S., et al., *N-glycosylation Profiling of Colorectal Cancer Cell Lines Reveals Association of Fucosylation with Differentiation and Caudal Type Homebox 1 (CDX1)/Villin mRNA Expression*.* Molecular & Cellular Proteomics, 2016. **15**(1): p. 124-140.
67. Ercan, A., et al., *Aberrant IgG galactosylation precedes disease onset, correlates with disease activity, and is prevalent in autoantibodies in rheumatoid arthritis.* Arthritis & Rheumatism, 2010. **62**(8): p. 2239-2248.
68. Lundström, S.L., et al., *IgG Fc galactosylation predicts response to methotrexate in early rheumatoid arthritis.* Arthritis Research & Therapy, 2017. **19**(1): p. 182.
69. Kemna, M.J., et al., *Galactosylation and Sialylation Levels of IgG Predict Relapse in Patients With PR3-ANCA Associated Vasculitis.* EBioMedicine, 2017. **17**: p. 108-118.

70. Rombouts, Y., et al., *Anti-citrullinated protein antibodies acquire a pro-inflammatory Fc glycosylation phenotype prior to the onset of rheumatoid arthritis.* Annals of the Rheumatic Diseases, 2015. **74**(1): p. 234-241.
71. McGarrah, R.W., et al., *A Novel Protein Glycan–Derived Inflammation Biomarker Independently Predicts Cardiovascular Disease and Modifies the Association of HDL Subclasses with Mortality.* Clinical Chemistry, 2017. **63**(1): p. 288-296.
72. Gudelj, I., et al., *Low galactosylation of IgG associates with higher risk for future diagnosis of rheumatoid arthritis during 10 years of follow-up.* Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease, 2018. **1864**(6, Part A): p. 2034-2039.
73. Cerf, M.E., *Transcription factors regulating β-cell function.* European Journal of Endocrinology, 2006. **155**(5): p. 671-679.
74. Cardenas-Diaz, F.L., et al., *Modeling Monogenic Diabetes using Human ESCs Reveals Developmental and Metabolic Deficiencies Caused by Mutations in HNF1A.* Cell Stem Cell, 2019. **25**(2): p. 273-289.e5.
75. Harries, L.W., et al., *Isomers of the TCF1 gene encoding hepatocyte nuclear factor-1 alpha show differential expression in the pancreas and define the relationship between mutation position and clinical phenotype in monogenic diabetes.* Human Molecular Genetics, 2006. **15**(14): p. 2216-2224.
76. Byrne, M.M., et al., *Altered Insulin Secretory Responses to Glucose in Diabetic and Nondiabetic Subjects With Mutations in the Diabetes Susceptibility Gene MODY3 on Chromosome 12.* Diabetes, 1996. **45**(11): p. 1503-1510.
77. Steele, A.M., et al., *Increased all-cause and cardiovascular mortality in monogenic diabetes as a result of mutations in the HNF1A gene.* Diabetic Medicine, 2010. **27**(2): p. 157-161.
78. Shields, B.M., et al., *The development and validation of a clinical prediction model to determine the probability of MODY in patients with young-onset diabetes.* Diabetologia, 2012. **55**(5): p. 1265-1272.
79. Shields, B.M., et al., *Maturity-onset diabetes of the young (MODY): how many cases are we missing?* Diabetologia, 2010. **53**(12): p. 2504-2508.
80. Najmi, L.A., et al., *Functional Investigations of HNF1A Identify Rare Variants as Risk Factors for Type 2 Diabetes in the General Population.* Diabetes, 2016. **66**(2): p. 335-346.
81. Urakami, T., *Maturity-onset diabetes of the young (MODY): current perspectives on diagnosis and treatment.* Diabetes Metab Syndr Obes, 2019. **12**: p. 1047-1056.
82. 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 Genet, 2010. **6**(12): p. e1001256.
83. Ventham, N.T., et al., *Changes to serum sample tube and processing methodology does not cause Intra-Individual [corrected] variation in automated whole serum N-glycan profiling in health and disease.* PLoS One, 2015. **10**(4): p. e0123028.
84. Kozak, R.P., et al., *Comparison of procainamide and 2-aminobenzamide labeling for profiling and identification of glycans by liquid chromatography with fluorescence detection coupled to electrospray ionization–mass spectrometry.* Analytical Biochemistry, 2015. **486**: p. 38-40.
85. Wu, H., et al., *Fucosidases from the human gut symbiont Ruminococcus gnavus.* Cellular and Molecular Life Sciences, 2021. **78**(2): p. 675-693.
86. O'Flaherty, R., et al., *Aminoquinoline Fluorescent Labels Obstruct Efficient Removal of N-Glycan Core α(1–6) Fucose by Bovine Kidney α-l-Fucosidase (BKF).* Journal of Proteome Research, 2017. **16**(11): p. 4237-4243.
87. Saldova, R., et al., *Association of N-Glycosylation with Breast Carcinoma and Systemic Features Using High-Resolution Quantitative UPLC.* Journal of Proteome Research, 2014. **13**(5): p. 2314-2327.
88. Jansen, B.C., et al., *HappyTools: A software for high-throughput HPLC data processing and quantitation.* PLoS One, 2018. **13**(7): p. e0200280.
89. Kwak, S.K. and J.H. Kim, *Statistical data preparation: management of missing values and outliers.* Korean J Anesthesiol, 2017. **70**(4): p. 407-411.

90. Gumustas, M., et al., *The History of the Core–Shell Particles and Applications in Active Pharmaceutical Ingredients Via Liquid Chromatography.* Chromatographia, 2019. **82**(1): p. 17-48.
91. Mucha, E., et al., *Fucose Migration in Intact Protonated Glycan Ions: A Universal Phenomenon in Mass Spectrometry.* Angewandte Chemie International Edition, 2018. **57**(25): p. 7440-7443.
92. Wuhrer, M., et al., *Mass spectrometry of proton adducts of fucosylated N-glycans: fucose transfer between antennae gives rise to misleading fragments.* Rapid Communications in Mass Spectrometry, 2006. **20**(11): p. 1747-1754.
93. Fournier, T., N. Medjoubi-N, and D. Porquet, *Alpha-1-acid glycoprotein.* Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology, 2000. **1482**(1): p. 157-171.
94. Clerc, F., et al., *Human plasma protein N-glycosylation.* Glycoconjugate Journal, 2016. **33**(3): p. 309-343.
95. Ayoya, M.A., et al., *α1-Acid glycoprotein, hepcidin, C-reactive protein, and serum ferritin are correlated in anemic schoolchildren with Schistosoma haematobium123.* The American Journal of Clinical Nutrition, 2010. **91**(6): p. 1784-1790.
96. Higai, K., et al., *Glycosylation of site-specific glycans of α1-acid glycoprotein and alterations in acute and chronic inflammation.* Biochimica et Biophysica Acta (BBA) - General Subjects, 2005. **1725**(1): p. 128-135.
97. Rudman, N., O. Gornik, and G. Lauc, *Altered N-glycosylation profiles as potential biomarkers and drug targets in diabetes.* FEBS Letters, 2019. **593**(13): p. 1598-1615.
98. Ogawa, K., et al., *Tri-antennary tri-sialylated mono-fucosylated glycan of alpha-1 antitrypsin as a non-invasive biomarker for non-alcoholic steatohepatitis: a novel glycobiomarker for non-alcoholic steatohepatitis.* Scientific Reports, 2020. **10**(1): p. 321.
99. Lee, J., et al., *High-normal levels of hs-CRP predict the development of non-alcoholic fatty liver in healthy men.* PLoS One, 2017. **12**(2): p. e0172666.
100. Kailemia, M.J., D. Park, and C.B. Lebrilla, *Glycans and glycoproteins as specific biomarkers for cancer.* Analytical and Bioanalytical Chemistry, 2017. **409**(2): p. 395-410.
101. Varki, A., *Biological roles of glycans.* Glycobiology, 2017. **27**(1): p. 3-49.
102. Knezevic, A., et al., *Effects of aging, body mass index, plasma lipid profiles, and smoking on human plasma N-glycans.* Glycobiology, 2010. **20**(8): p. 959-69.
103. Gudelj, I., et al., *Low galactosylation of IgG associates with higher risk for future diagnosis of rheumatoid arthritis during 10 years of follow-up.* Biochim Biophys Acta Mol Basis Dis, 2018. **1864**(6 Pt A): p. 2034-2039.
104. Ercan, A., et al., *Aberrant IgG galactosylation precedes disease onset, correlates with disease activity, and is prevalent in autoantibodies in rheumatoid arthritis.* Arthritis Rheum, 2010. **62**(8): p. 2239-48.
105. Gudelj, I., et al., *Estimation of human age using N-glycan profiles from bloodstains.* International Journal of Legal Medicine, 2015. **129**(5): p. 955-961.
106. Pearson, E.R., et al., *Genetic cause of hyperglycaemia and response to treatment in diabetes.* The Lancet, 2003. **362**(9392): p. 1275-1281.
107. Leeman, M., et al., *Proteins and antibodies in serum, plasma, and whole blood—size characterization using asymmetrical flow field-flow fractionation (AF4).* Analytical and Bioanalytical Chemistry, 2018. **410**(20): p. 4867-4873.
108. Shields, B.M., et al., *Population-Based Assessment of a Biomarker-Based Screening Pathway to Aid Diagnosis of Monogenic Diabetes in Young-Onset Patients.* Diabetes Care, 2017. **40**(8): p. 1017-1025.
109. Misra, S. and K.R. Owen, *Genetics of Monogenic Diabetes: Present Clinical Challenges.* Current Diabetes Reports, 2018. **18**(12): p. 141.
110. Becker, D.J. and J.B. Lowe, *Fucose: biosynthesis and biological function in mammals.* Glycobiology, 2003. **13**(7): p. 41R-53R.
111. Ellulu, M.S., et al., *Obesity and inflammation: the linking mechanism and the complications.* Arch Med Sci, 2017. **13**(4): p. 851-863.
112. Ding, N., et al., *Human serum N-glycan profiles are age and sex dependent.* Age and Ageing, 2011. **40**(5): p. 568-575.

113. Young, B., M. Gleeson, and A.W. Cripps, *C-reactive protein: A critical review.* Pathology, 1991. **23**(2): p. 118-124.
114. de Vries, T., et al., *Fucosyltransferases: structure/function studies.* Glycobiology, 2001. **11**(10): p. 119R-128R.
115. King, J.R., J. Varadé, and L. Hammarström, *Fucosyltransferase Gene Polymorphisms and Lewisb-Negative Status Are Frequent in Swedish Newborns, With Implications for Infectious Disease Susceptibility and Personalized Medicine.* Journal of the Pediatric Infectious Diseases Society, 2019. **8**(6): p. 507-518.
116. Haab, B.B., *Using lectins in biomarker research: addressing the limitations of sensitivity and availability.* Proteomics Clin Appl, 2012. **6**(7-8): p. 346-50.
117. Mahley, R.W., et al., *Plasma lipoproteins: apolipoprotein structure and function.* Journal of Lipid Research, 1984. **25**(12): p. 1277-1294.
118. Shachter, N.S., *Apolipoproteins C-I and C-III as important modulators of lipoprotein metabolism.* Current Opinion in Lipidology, 2001. **12**(3).
119. Larsson, M., et al., *Apolipoprotein C-III inhibits triglyceride hydrolysis by GPIHBP1-bound LPL[S].* Journal of Lipid Research, 2017. **58**(9): p. 1893-1902.
120. Larsson, M., et al., *Apolipoproteins C-I and C-III Inhibit Lipoprotein Lipase Activity by Displacement of the Enzyme from Lipid Droplets*.* Journal of Biological Chemistry, 2013. **288**(47): p. 33997-34008.
121. Kawakami, A., et al., *Apolipoprotein CIII in Apolipoprotein B Lipoproteins Enhances the Adhesion of Human Monocytic Cells to Endothelial Cells.* Circulation, 2006. **113**(5): p. 691-700.
122. Olin-Lewis, K., et al., *ApoC-III content of apoB-containing lipoproteins is associated with binding to the vascular proteoglycan biglycan.* Journal of Lipid Research, 2002. **43**(11): p. 1969-1977.
123. Dai, W., et al., *Emerging evidences for the opposite role of apolipoprotein C3 and apolipoprotein A5 in lipid metabolism and coronary artery disease.* Lipids in Health and Disease, 2019. **18**(1): p. 220.
124. Kohan, A.B., *Apolipoprotein C-III: a potent modulator of hypertriglyceridemia and cardiovascular disease.* Current Opinion in Endocrinology, Diabetes and Obesity, 2015. **22**(2).
125. Taskinen, M.-R., C.J. Packard, and J. Borén, *Emerging Evidence that ApoC-III Inhibitors Provide Novel Options to Reduce the Residual CVD.* Current Atherosclerosis Reports, 2019. **21**(8): p. 27.
126. Rosenson, R.S., et al., *Dysfunctional HDL and atherosclerotic cardiovascular disease.* Nature Reviews Cardiology, 2016. **13**(1): p. 48-60.
127. Wyler von Ballmoos, M.C., B. Haring, and F.M. Sacks, *The risk of cardiovascular events with increased apolipoprotein CIII: A systematic review and meta-analysis.* Journal of Clinical Lipidology, 2015. **9**(4): p. 498-510.
128. Christopoulou, E., et al., *Apolipoprotein CIII and diabetes. Is there a link?* Diabetes/Metabolism Research and Reviews, 2019. **35**(3): p. e3118.
129. Juntti-Berggren, L. and P.-O. Berggren, *Apolipoprotein CIII is a new player in diabetes.* Current Opinion in Lipidology, 2017. **28**(1).
130. Pollin, T.I., et al., *A Null Mutation in Human APOC3 Confers a Favorable Plasma Lipid Profile and Apparent Cardioprotection.* Science, 2008. **322**(5908): p. 1702-1705.
131. Rocha, N.A., et al., *ApoCIII as a Cardiovascular Risk Factor and Modulation by the Novel Lipid-Lowering Agent Volanesorsen.* Current Atherosclerosis Reports, 2017. **19**(12): p. 62.
132. Reyes-Soffer, G., et al., *Effects of APOC3 Heterozygous Deficiency on Plasma Lipid and Lipoprotein Metabolism.* Arteriosclerosis, Thrombosis, and Vascular Biology, 2019. **39**(1): p. 63-72.
133. Nedelkov, D. *Mass Spectrometric Studies of Apolipoprotein Proteoforms and Their Role in Lipid Metabolism and Type 2 Diabetes*. Proteomes, 2017. **5**, DOI: 10.3390/proteomes5040027.
134. Ramms, B. and P.L.S.M. Gordts, *Apolipoprotein C-III in triglyceride-rich lipoprotein metabolism.* Current Opinion in Lipidology, 2018. **29**(3).
135. Holleboom, A.G., et al., *Heterozygosity for a loss-of-function mutation in GALNT2 improves plasma triglyceride clearance in man.* Cell Metab, 2011. **14**(6): p. 811-8.
136. Hiukka, A., et al., *ApoCIII-enriched LDL in type 2 diabetes displays altered lipid composition, increased susceptibility for sphingomyelinase, and increased binding to biglycan.* Diabetes, 2009. **58**(9): p. 2018-26.

137. Nicolardi, S., et al., *Mapping O-glycosylation of apolipoprotein C-III in MALDI-FT-ICR protein profiles.* Proteomics, 2013. **13**(6): p. 992-1001.
138. Stadtman, E.R., J. Moskovitz, and R.L. Levine, *Oxidation of Methionine Residues of Proteins: Biological Consequences.* Antioxidants & Redox Signaling, 2003. **5**(5): p. 577-582.
139. Borges, C.R., et al., *Elevated Plasma Albumin and Apolipoprotein A-I Oxidation under Suboptimal Specimen Storage Conditions*.* Molecular & Cellular Proteomics, 2014. **13**(7): p. 1890-1899.
140. Jansen, B.C., et al., *MassyTools: A High-Throughput Targeted Data Processing Tool for Relative Quantitation and Quality Control Developed for Glycomic and Glycoproteomic MALDI-MS.* Journal of Proteome Research, 2015. **14**(12): p. 5088-5098.
141. van Herpt, T.T.W., et al., *Introduction of the DiaGene study: clinical characteristics, pathophysiology and determinants of vascular complications of type 2 diabetes.* Diabetol Metab Syndr, 2017. **9**: p. 47.
142. van der Burgt, Y.E.M., et al., *Structural Analysis of Monoclonal Antibodies by Ultrahigh Resolution MALDI In-Source Decay FT-ICR Mass Spectrometry.* Analytical Chemistry, 2019. **91**(3): p. 2079-2085.
143. Lao, Y.W., et al., *Chromatographic behavior of peptides containing oxidized methionine residues in proteomic LC–MS experiments: Complex tale of a simple modification.* Journal of Proteomics, 2015. **125**: p. 131-139.
144. Hains, P.G. and P.J. Robinson, *The Impact of Commonly Used Alkylating Agents on Artifactual Peptide Modification.* Journal of Proteome Research, 2017. **16**(9): p. 3443-3447.
145. Bettinger, J.Q., et al., *Quantitative Analysis of in Vivo Methionine Oxidation of the Human Proteome.* Journal of Proteome Research, 2020. **19**(2): p. 624-633.
146. Kim, G., S.J. Weiss, and R.L. Levine, *Methionine oxidation and reduction in proteins.* Biochimica et Biophysica Acta (BBA) - General Subjects, 2014. **1840**(2): p. 901-905.
147. Lim, J.M., G. Kim, and R.L. Levine, *Methionine in Proteins: It's Not Just for Protein Initiation Anymore.* Neurochemical Research, 2019. **44**(1): p. 247-257.
148. Trenchevska, O., R.W. Nelson, and D. Nedelkov *Mass Spectrometric Immunoassays in Characterization of Clinically Significant Proteoforms*. Proteomes, 2016. **4**, DOI: 10.3390/proteomes4010013.
149. Kailemia, M.J., et al., *Targeted Measurements of O- and N-Glycopeptides Show That Proteins in High Density Lipoprotein Particles Are Enriched with Specific Glycosylation Compared to Plasma.* Journal of Proteome Research, 2018. **17**(2): p. 834-845.
150. Olivieri, O., et al., *Sialylated isoforms of apolipoprotein C-III and plasma lipids in subjects with coronary artery disease.* 2018. **56**(9): p. 1542-1550.
151. Nelsestuen, G.L., et al., *Plasma protein profiling: unique and stable features of individuals.* Proteomics, 2005. **5**(15): p. 4012-24.
152. Palmigiano, A., et al., *MALDI-MS profiling of serum O-glycosylation and N-glycosylation in COG5-CDG.* Journal of Mass Spectrometry, 2017. **52**(6): p. 372-377.
153. Wada, Y. and N. Okamoto, *Apolipoprotein C-III O-glycoform profiling of 500 serum samples by matrix-assisted laser desorption/ionization mass spectrometry for diagnosis of congenital disorders of glycosylation.* Journal of Mass Spectrometry, 2021. **56**(4): p. e4597.
154. Nicolardi, S., et al., *Quality control based on isotopic distributions for high-throughput MALDI-TOF and MALDI-FTICR serum peptide profiling.* Journal of the American Society for Mass Spectrometry, 2010. **21**(9): p. 1515-1525.
155. Oldoni, F., et al., *Naturally Occurring Variants in LRP1 (Low-Density Lipoprotein Receptor–Related Protein 1) Affect HDL (High-Density Lipoprotein) Metabolism Through ABCA1 (ATP-Binding Cassette A1) and SR-B1 (Scavenger Receptor Class B Type 1) in Humans.* Arteriosclerosis, Thrombosis, and Vascular Biology, 2018. **38**(7): p. 1440-1453.
156. Paththinige, C.S., et al., *Spectrum of low-density lipoprotein receptor (LDLR) mutations in a cohort of Sri Lankan patients with familial hypercholesterolemia – a preliminary report.* Lipids in Health and Disease, 2018. **17**(1): p. 100.
157. Xu, Y.-X., et al., *EDEM3 Modulates Plasma Triglyceride Level through Its Regulation of LRP1 Expression.* iScience, 2020. **23**(4): p. 100973.

158. Kim, H.J., et al., *The hypolipidemic effect of cilostazol can be mediated by regulation of hepatic low-density lipoprotein receptor-related protein 1 (LRP1) expression.* Metabolism, 2014. **63**(1): p. 112-119.
159. Franssen, R., et al., *Obesity and Dyslipidemia.* Medical Clinics of North America, 2011. **95**(5): p. 893-902.
160. Borén, J., C.J. Packard, and M.-R. Taskinen, *The Roles of ApoC-III on the Metabolism of Triglyceride-Rich Lipoproteins in Humans.* Frontiers in Endocrinology, 2020. **11**.
161. Mauger, J.F., et al., *Apolipoprotein C-III isoforms: kinetics and relative implication in lipid metabolism.* J Lipid Res, 2006. **47**(6): p. 1212-8.
162. Kyu, H.H., et al., *Global, regional, and national disability-adjusted life-years (DALYs) for 359 diseases and injuries and healthy life expectancy (HALE) for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017.* The Lancet, 2018. **392**(10159): p. 1859-1922.
163. Sone, H., et al., *Serum Level of Triglycerides Is a Potent Risk Factor Comparable to LDL Cholesterol for Coronary Heart Disease in Japanese Patients with Type 2 Diabetes: Subanalysis of the Japan Diabetes Complications Study (JDCS).* The Journal of Clinical Endocrinology & Metabolism, 2011. **96**(11): p. 3448-3456.
164. Ginsberg, H.N., et al., *Apolipoprotein B metabolism in subjects with deficiency of apolipoproteins CIII and AI. Evidence that apolipoprotein CIII inhibits catabolism of triglyceride-rich lipoproteins by lipoprotein lipase in vivo.* The Journal of Clinical Investigation, 1986. **78**(5): p. 1287-1295.
165. Kinnunen, P.K.J. and C. Ehnholm, *Effect of serum and C-apoproteins from very low density lipoproteins on human postheparin plasma hepatic lipase.* FEBS Letters, 1976. **65**(3): p. 354-357.
166. Zhang, J., N.A. Rocha, and P.A. McCullough, *Contribution of ApoCIII to Diabetic Dyslipidemia and Treatment With Volanesorsen.* RCM, 2018. **19**(1): p. 32-37.
167. Sacks, E.B.F.M., C. Zheng, and E.B.J.S. Cohn, *Complexities of plasma apolipoprotein C-III metabolism1.* Journal of Lipid Research, 2011. **52**(6): p. 1067-1070.
168. Fredenrich, A., et al., *Plasma lipoprotein distribution of apoC-III in normolipidemic and hypertriglyceridemic subjects: comparison of the apoC-III to apoE ratio in different lipoprotein fractions.* Journal of Lipid Research, 1997. **38**(7): p. 1421-1432.
169. Valladolid-Acebes, I., P.-O. Berggren, and L. Juntti-Berggren *Apolipoprotein CIII Is an Important Piece in the Type-1 Diabetes Jigsaw Puzzle*. International Journal of Molecular Sciences, 2021. **22**, DOI: 10.3390/ijms22020932.
170. Jørgensen, A.B., et al., *Loss-of-Function Mutations in APOC3 and Risk of Ischemic Vascular Disease.* New England Journal of Medicine, 2014. **371**(1): p. 32-41.
171. Kawakami, A., et al., *Apolipoprotein CIII Induces Expression of Vascular Cell Adhesion Molecule-1 in Vascular Endothelial Cells and Increases Adhesion of Monocytic Cells.* Circulation, 2006. **114**(7): p. 681-687.
172. Juntti-Berggren, L., et al., *Apolipoprotein CIII promotes Ca2+-dependent β cell death in type 1 diabetes.* Proceedings of the National Academy of Sciences, 2004. **101**(27): p. 10090-10094.
173. Digenio, A., et al., *Antisense-Mediated Lowering of Plasma Apolipoprotein C-III by Volanesorsen Improves Dyslipidemia and Insulin Sensitivity in Type 2 Diabetes.* Diabetes Care, 2016. **39**(8): p. 1408-1415.
174. van Hoek, M., et al., *Association of an APOC3 promoter variant with type 2 diabetes risk and need for insulin treatment in lean persons.* Diabetologia, 2011. **54**(6): p. 1360-1367.
175. Schjoldager, K.T.B.G., et al., *Probing isoform-specific functions of polypeptide GalNAc-transferases using zinc finger nuclease glycoengineered SimpleCells.* Proceedings of the National Academy of Sciences, 2012. **109**(25): p. 9893-9898.
176. Surakka, I., et al., *The impact of low-frequency and rare variants on lipid levels.* Nature Genetics, 2015. **47**(6): p. 589-597.
177. Santos Seckler, H.d., et al., *New Interface for Faster Proteoform Analysis: Immunoprecipitation Coupled with SampleStream-Mass Spectrometry.* Journal of the American Society for Mass Spectrometry, 2021. **32**(7): p. 1659-1670.
178. Holleboom, Adriaan G., et al., *Heterozygosity for a Loss-of-Function Mutation in GALNT2 Improves Plasma Triglyceride Clearance in Man.* Cell Metabolism, 2011. **14**(6): p. 811-818.

179. Demus, D., et al., *Large-Scale Analysis of Apolipoprotein CIII Glycosylation by Ultrahigh Resolution Mass Spectrometry.* Front Chem, 2021. **9**: p. 678883.
180. van der Heijden, A.A., et al., *The Hoorn Diabetes Care System (DCS) cohort. A prospective cohort of persons with type 2 diabetes treated in primary care in the Netherlands.* BMJ Open, 2017. **7**(5): p. e015599.
181. Purcell, S., et al., *PLINK: a tool set for whole-genome association and population-based linkage analyses.* Am J Hum Genet, 2007. **81**(3): p. 559-75.
182. McCarthy, S., et al., *A reference panel of 64,976 haplotypes for genotype imputation.* Nat Genet, 2016. **48**(10): p. 1279-83.
183. Loh, P.R., et al., *Reference-based phasing using the Haplotype Reference Consortium panel.* Nat Genet, 2016. **48**(11): p. 1443-1448.
184. Das, S., et al., *Next-generation genotype imputation service and methods.* Nat Genet, 2016. **48**(10): p. 1284-1287.
185. Guo, T., et al., *Association between the DOCK7, PCSK9 and GALNT2 Gene Polymorphisms and Serum Lipid levels.* Scientific Reports, 2016. **6**(1): p. 19079.
186. Watanabe, K., et al., *Functional mapping and annotation of genetic associations with FUMA.* Nature Communications, 2017. **8**(1): p. 1826.
187. Zhan, X., et al., *RVTESTS: an efficient and comprehensive tool for rare variant association analysis using sequence data.* Bioinformatics, 2016. **32**(9): p. 1423-6.
188. Winkler, T.W., et al., *Quality control and conduct of genome-wide association meta-analyses.* Nat Protoc, 2014. **9**(5): p. 1192-212.
189. Type 2 Diabetes Knowledge Portal., n.d.r., rs67086575, and rs10842926 variant pages, *Type 2 Diabetes Knowledge Portal., n.d. rs4846913, rs67086575, and rs10842926 variant pages*.
190. Ben, E., et al., *The MRC IEU OpenGWAS data infrastructure.* bioRxiv, 2020: p. 2020.08.10.244293.
191. Võsa, U., et al., *Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression.* Nature Genetics, 2021. **53**(9): p. 1300-1310.
192. van den Boogert, M.A.W., D.J. Rader, and A.G. Holleboom, *New insights into the role of glycosylation in lipoprotein metabolism.* Current Opinion in Lipidology, 2017. **28**(6).
193. Schjoldager, K.T.B.G., et al., *O-Glycosylation Modulates Proprotein Convertase Activation of Angiopoietin-like Protein 3: POSSIBLE ROLE OF POLYPEPTIDE GalNAc-TRANSFERASE-2 IN REGULATION OF CONCENTRATIONS OF PLASMA LIPIDS*.* Journal of Biological Chemistry, 2010. **285**(47): p. 36293-36303.
194. Roman, Tamara S., et al., *Multiple Hepatic Regulatory Variants at the GALNT2 GWAS Locus Associated with High-Density Lipoprotein Cholesterol.* The American Journal of Human Genetics, 2015. **97**(6): p. 801-815.
195. Cavalli, M., et al., *Looking beyond GWAS: allele-specific transcription factor binding drives the association of GALNT2 to HDL-C plasma levels.* Lipids in Health and Disease, 2016. **15**(1): p. 18.
196. Guo, T., et al., *Association between the DOCK7, PCSK9 and GALNT2 Gene Polymorphisms and Serum Lipid levels.* Sci Rep, 2016. **6**: p. 19079.
197. Ridker, P.M., et al., *Polymorphism in the CETP gene region, HDL cholesterol, and risk of future myocardial infarction: Genomewide analysis among 18 245 initially healthy women from the Women's Genome Health Study.* Circ Cardiovasc Genet, 2009. **2**(1): p. 26-33.
198. Schjoldager, K.T., et al., *O-glycosylation modulates proprotein convertase activation of angiopoietin-like protein 3: possible role of polypeptide GalNAc-transferase-2 in regulation of concentrations of plasma lipids.* J Biol Chem, 2010. **285**(47): p. 36293-303.
199. Taschner, M., S. Bhogaraju, and E. Lorentzen, *Architecture and function of IFT complex proteins in ciliogenesis.* Differentiation, 2012. **83**(2): p. S12-22.
200. Võsa, U., et al., *Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression.* Nat Genet, 2021. **53**(9): p. 1300-1310.
201. De Langhe, S., et al., *Interaction of the small GTPase Rac3 with NRBP, a protein with a kinase-homology domain.* Int J Mol Med, 2002. **9**(5): p. 451-9.
202. Harduin-Lepers, A., et al., *The human sialyltransferase family.* Biochimie, 2001. **83**(8): p. 727-37.
203. Rodríguez, M., et al., *Distribution of seven ApoC-III glycoforms in plasma, VLDL, IDL, LDL and HDL of healthy subjects.* J Proteomics, 2022. **251**: p. 104398.

204. Organisation, W.H. *WHO factsheet Diabetes [Internet]*. 2020; Available from: https://www.who.int/news-room/fact-sheets/detail/diabetes.
205. Patel, K.V. and M. Vaduganathan, *Targeting multiple domains of residual cardiovascular disease risk in patients with diabetes.* Curr Opin Cardiol, 2020. **35**(5): p. 517-523.
206. Gaede, P., et al., *Multifactorial intervention and cardiovascular disease in patients with type 2 diabetes.* N Engl J Med, 2003. **348**(5): p. 383-93.
207. Chait, A., et al., *Remnants of the Triglyceride-Rich Lipoproteins, Diabetes, and Cardiovascular Disease.* Diabetes, 2020. **69**(4): p. 508-516.
208. Ginsberg, H.N., *Lipoprotein physiology in nondiabetic and diabetic states. Relationship to atherogenesis.* Diabetes Care, 1991. **14**(9): p. 839-55.
209. Norata, G.D., et al., *Apolipoprotein C-III: From Pathophysiology to Pharmacology.* Trends Pharmacol Sci, 2015. **36**(10): p. 675-687.
210. Kawakami, A. and M. Yoshida, *Apolipoprotein CIII links dyslipidemia with atherosclerosis.* J Atheroscler Thromb, 2009. **16**(1): p. 6-11.
211. Sundsten, T., C.G. Ostenson, and P. Bergsten, *Serum protein patterns in newly diagnosed type 2 diabetes mellitus--influence of diabetic environment and family history of diabetes.* Diabetes Metab Res Rev, 2008. **24**(2): p. 148-54.
212. Hiukka, A., et al., *Alterations of lipids and apolipoprotein CIII in very low density lipoprotein subspecies in type 2 diabetes.* Diabetologia, 2005. **48**(6): p. 1207-15.
213. Klein, R.L., et al., *Apolipoprotein C-III protein concentrations and gene polymorphisms in Type 1 diabetes: associations with microvascular disease complications in the DCCT/EDIC cohort.* J Diabetes Complications, 2005. **19**(1): p. 18-25.
214. Hu, Z.J., et al., *Associations between apolipoprotein CIII concentrations and microalbuminuria in type 2 diabetes.* Exp Ther Med, 2014. **8**(3): p. 951-956.
215. Zhang, Q., et al., *Relationship between serum apolipoproteins levels and retinopathy risk in subjects with type 2 diabetes mellitus.* Acta Diabetol, 2018. **55**(7): p. 681-689.
216. Jia, X., et al., *Highlights from Studies in Cardiovascular Disease Prevention Presented at the Digital 2020 European Society of Cardiology Congress: Prevention Is Alive and Well.* Curr Atheroscler Rep, 2020. **22**(12): p. 72.
217. Nurmohamed, N.S., G.M. Dallinga-Thie, and E.S.G. Stroes, *Targeting apoC-III and ANGPTL3 in the treatment of hypertriglyceridemia.* Expert Rev Cardiovasc Ther, 2020. **18**(6): p. 355-361.
218. Mendoza, S., et al., *Changes in low-density lipoprotein size phenotypes associate with changes in apolipoprotein C-III glycoforms after dietary interventions.* J Clin Lipidol, 2017. **11**(1): p. 224-233.e2.
219. American Diabetes, A., *Diagnosis and Classification of Diabetes Mellitus.* Diabetes Care, 2013. **37**(Supplement_1): p. S81-S90.
220. Alberti, K.G. and P.Z. Zimmet, *Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: diagnosis and classification of diabetes mellitus provisional report of a WHO consultation.* Diabet Med, 1998. **15**(7): p. 539-53.
221. Naber, A., et al., *Apolipoprotein-CIII O-Glycosylation, a Link between GALNT2 and Plasma Lipids.* International Journal of Molecular Sciences, 2023. **24**(19): p. 14844.
222. Schwarzer, G. *Package "meta", Version 7.0-0*. 2024 15 March 2024]; Available from: https://cran.r-project.org/web/packages/meta/index.html.
223. Higgins, J.P., et al., *Measuring inconsistency in meta-analyses.* Bmj, 2003. **327**(7414): p. 557-60.
224. Caron, S., et al., *Transcriptional activation of apolipoprotein CIII expression by glucose may contribute to diabetic dyslipidemia.* Arterioscler Thromb Vasc Biol, 2011. **31**(3): p. 513-9.
225. Adiels, M., et al., *Role of apolipoprotein C-III overproduction in diabetic dyslipidaemia.* Diabetes Obes Metab, 2019. **21**(8): p. 1861-1870.
226. Geng, T., et al., *Healthy lifestyle behaviors, mediating biomarkers, and risk of microvascular complications among individuals with type 2 diabetes: A cohort study.* PLoS Med, 2023. **20**(1): p. e1004135.
227. Sillanaukee, P., M. Pönniö, and I.P. Jääskeläinen, *Occurrence of sialic acids in healthy humans and different disorders.* Eur J Clin Invest, 1999. **29**(5): p. 413-25.

228. Watts, G.F., et al., *Serum sialic acid as an indicator of change in coronary artery disease.* Metabolism, 1995. **44**(2): p. 147-8.
229. Harake, B., et al., *A simple micromethod for rapid assessment of the distribution of apolipoprotein C isoforms in very-low-density lipoprotein.* Clin Biochem, 1991. **24**(3): p. 255-60.
230. Rodríguez, M., et al., *Distribution of seven ApoC-III glycoforms in plasma, VLDL, IDL, LDL and HDL of healthy subjects.* Journal of Proteomics, 2022. **251**: p. 104398.
231. *ICH Q6B Specifications: test procedures and acceptance criteria for biotechnological/biological products - Scientific guideline* 2023 05 May 2023]; Available from: https://www.ema.europa.eu/en/documents/scientific-guideline/ich-q-6-b-test-procedures-acceptance-criteria-biotechnological/biological-products-step-5_en.pdf.
232. *Guidance for Industry PAT — A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance*. 2023 [cited 2023; Available from: https://www.fda.gov/media/71012/download.
233. *ICH, Q5E Specifications: Test Procedures and Acceptance Criteria for Biotechnological/Biological Products, EMA Document CPMP/ICH/5721/03 (Geneva, 2003)* 2023 [cited 2023; Available from: https://www.ema.europa.eu/en/documents/scientific-guideline/ich-q-5-e-comparability-biotechnological/biological-products-step-5_en.pdf.
234. Byeon, S.K., et al., *Development of a multiomics model for identification of predictive biomarkers for COVID-19 severity: a retrospective cohort study.* Lancet Digit Health, 2022. **4**(9): p. e632-e645.
235. Agrawal, M., et al., *Multiomics to elucidate inflammatory bowel disease risk factors and pathways.* Nat Rev Gastroenterol Hepatol, 2022. **19**(6): p. 399-409.
236. Rudman, N., et al., *Integrated glycomics and genetics analyses reveal a potential role for N-glycosylation of plasma proteins and IgGs, as well as the complement system, in the development of type 1 diabetes.* Diabetologia, 2023. **66**(6): p. 1071-1083.
237. Luo, X.J., et al., *Novel Genetic and Epigenetic Biomarkers of Prognostic and Predictive Significance in Stage II/III Colorectal Cancer.* Mol Ther, 2021. **29**(2): p. 587-596.
238. Shubhakar, A., et al., *Serum N-Glycomic Biomarkers Predict Treatment Escalation in Inflammatory Bowel Disease.* J Crohns Colitis, 2023. **17**(6): p. 919-932.
239. Ansari, D., et al., *Proteomic and genomic profiling of pancreatic cancer.* Cell Biol Toxicol, 2019. **35**(4): p. 333-343.
240. Adams, S.A. and C. Petersen, *Precision medicine: opportunities, possibilities, and challenges for patients and providers.* Journal of the American Medical Informatics Association, 2016. **23**(4): p. 787-790.
241. Misra, S. and K.R. Owen, *Genetics of Monogenic Diabetes: Present Clinical Challenges.* Curr Diab Rep, 2018. **18**(12): p. 141.
242. *ISO - 03.100.70 - Management systems* [cited 2023; Available from: www.iso.org/ics/03.100.70/x/.
243. *ISO 9001:2015 Quality management systems — Requirements*. [cited 2023; Available from: www.iso.org/standard/62085.html.
244. Song, J.W. and K.C. Chung, *Observational studies: cohort and case-control studies.* Plast Reconstr Surg, 2010. **126**(6): p. 2234-2242.
245. Naber, A., et al., *Apolipoprotein-CIII O-Glycosylation Is Associated with Micro- and Macrovascular Complications of Type 2 Diabetes.* International Journal of Molecular Sciences, 2024. **25**(10): p. 5365.
246. Singh, S.S., et al., *Association of the IgG N-glycome with the course of kidney function in type 2 diabetes.* BMJ Open Diabetes Res Care, 2020. **8**(1).
247. Vreeker, G.C.M., et al., *Serum N-Glycome analysis reveals pancreatic cancer disease signatures.* Cancer Med, 2020. **9**(22): p. 8519-8529.
248. Amez Martín, M., M. Wuhrer, and D. Falck, *Serum and Plasma Immunoglobulin G Fc N-Glycosylation Is Stable during Storage.* Journal of Proteome Research, 2021. **20**(5): p. 2935-2941.
249. Torok, R., et al., *N-Glycosylation Profiling of Human Blood in Type 2 Diabetes by Capillary Electrophoresis: A Preliminary Study.* Molecules, 2021. **26**(21).
250. Mitchell, B.L., et al., *Impact of freeze-thaw cycles and storage time on plasma samples used in mass spectrometry based biomarker discovery projects.* Cancer Inform, 2005. **1**(1): p. 98-104.

251. Jian, W., et al., *Relative Quantitation of Glycoisoforms of Intact Apolipoprotein C3 in Human Plasma by Liquid Chromatography–High-Resolution Mass Spectrometry.* Analytical Chemistry, 2013. **85**(5): p. 2867-2874.
252. Trenchevska, O., et al., *Development of multiplex mass spectrometric immunoassay for detection and quantification of apolipoproteins C-I, C-II, C-III and their proteoforms.* Methods, 2015. **81**: p. 86-92.
253. Wada, Y., M. Kadoya, and N. Okamoto, *Mass spectrometry of apolipoprotein C-III, a simple analytical method for mucin-type O-glycosylation and its application to an autosomal recessive cutis laxa type-2 (ARCL2) patient.* Glycobiology, 2012. **22**(8): p. 1140-1144.
254. Yen-Nicolaÿ, S., et al., *MALDI-TOF MS applied to apoC-III glycoforms of patients with congenital disorders affecting O-glycosylation. Comparison with two-dimensional electrophoresis.* PROTEOMICS – Clinical Applications, 2015. **9**(7-8): p. 787-793.
255. Uffelmann, E., et al., *Genome-wide association studies.* Nature Reviews Methods Primers, 2021. **1**(1): p. 59.
256. Cano-Gamez, E. and G. Trynka, *From GWAS to Function: Using Functional Genomics to Identify the Mechanisms Underlying Complex Diseases.* Front Genet, 2020. **11**: p. 424.
257. Ben-Zeev, O., et al., *Lipoprotein lipase and hepatic lipase: the role of asparagine-linked glycosylation in the expression of a functional enzyme.* J Lipid Res, 1994. **35**(9): p. 1511-23.
258. Gibney, E.R. and C.M. Nolan, *Epigenetics and gene expression.* Heredity, 2010. **105**(1): p. 4-13.
259. Lacal, I. and R. Ventura, *Epigenetic Inheritance: Concepts, Mechanisms and Perspectives.* Frontiers in Molecular Neuroscience, 2018. **11**.
260. Ling, C., *Epigenetic regulation of insulin action and secretion – role in the pathogenesis of type 2 diabetes.* Journal of Internal Medicine, 2020. **288**(2): p. 158-167.
261. Bridgeman, S.C., et al., *Epigenetic effects of metformin: From molecular mechanisms to clinical implications.* Diabetes, Obesity and Metabolism, 2018. **20**(7): p. 1553-1562.
262. Ochoa-Rosales, C., et al., *Epigenetic Link Between Statin Therapy and Type 2 Diabetes.* Diabetes Care, 2020. **43**(4): p. 875-884.
263. Reiding, K.R., et al., *High-throughput Serum N-Glycomics: Method Comparison and Application to Study Rheumatoid Arthritis and Pregnancy-associated Changes*[S].* Molecular & Cellular Proteomics, 2019. **18**(1): p. 3-15.
264. Huffman, J.E., et al., *Comparative Performance of Four Methods for High-throughput Glycosylation Analysis of Immunoglobulin G in Genetic and Epidemiological Research*.* Molecular & Cellular Proteomics, 2014. **13**(6): p. 1598-1610.
265. Petretti, T., et al., *Altered mRNA expression of glycosyltransferases in human colorectal carcinomas and liver metastases.* Gut, 2000. **46**(3): p. 359-66.
266. Costa, A.F., et al., *Targeting Glycosylation: A New Road for Cancer Drug Discovery.* Trends Cancer, 2020. **6**(9): p. 757-766.
267. Fazal, F., et al., *The rising cost of healthcare and its contribution to the worsening disease burden in developing countries.* Annals of Medicine and Surgery, 2022. **82**.
268. Holman, H.R., *The Relation of the Chronic Disease Epidemic to the Health Care Crisis.* ACR Open Rheumatology, 2020. **2**(3): p. 167-173.
269. Edwards, E., et al., *Strategies to control therapeutic antibody glycosylation during bioprocessing: Synthesis and separation.* Biotechnology and Bioengineering, 2022. **119**(6): p. 1343-1358.
270. Li, S., et al. *Glycoengineering of Therapeutic Antibodies with Small Molecule Inhibitors*. Antibodies, 2021. **10**, DOI: 10.3390/antib10040044.
271. Fisher, P., et al., *The N-Glycosylation Processing Potential of the Mammalian Golgi Apparatus.* Frontiers in Cell and Developmental Biology, 2019. **7**.
272. Fisher, P., et al., *Modeling Glycan Processing Reveals Golgi-Enzyme Homeostasis upon Trafficking Defects and Cellular Differentiation.* Cell Reports, 2019. **27**(4): p. 1231-1243.e6.
273. Stork, R., et al., *N-Glycosylation as Novel Strategy to Improve Pharmacokinetic Properties of Bispecific Single-chain Diabodies*.* Journal of Biological Chemistry, 2008. **283**(12): p. 7804-7812.
274. Li, C., et al., *Protein Engineering for Improving and Diversifying Natural Product Biosynthesis.* Trends in Biotechnology, 2020. **38**(7): p. 729-744.

275. Mouchahoir, T. and J.E. Schiel, *Development of an LC-MS/MS peptide mapping protocol for the NISTmAb.* Anal Bioanal Chem, 2018. **410**(8): p. 2111-2126.

English Summary

The core of the research presented in this thesis lies in glycan biomarkers discovery in diabetes, analytical method development, and their clinical implications. Glycan biomarkers have emerged as promising tools for diagnosing, prognosing, and monitoring various diseases, including cancer, autoimmune diseases, and infectious diseases. This thesis focuses on developing and optimizing analytical methods to study glycosylation changes as potential biomarkers in large clinical cohorts of patients with diabetes and related complications.

The research presented aims to address challenges in high-throughput sample preparation of glycosylation analysis, data processing, and statistical analysis. The developed methods are applicable for analysing *N*-glycans, *O*-glycosylated proteins, and absolute fucosylation levels of proteins in blood plasma.

The thesis opens with a general introduction describing types of glycosylation, glyco(proteo)mics analytical approaches, and glycan biomarkers in diabetes (**Chapter 1**). The research part begins with an exploration of *N*-glycan antennary fucosylation as a biomarker for a monogenic type of diabetes, HNF1A-MODY (**Chapter 2**). Individuals with HNF1A-MODY carry variants in the *HNF1A* gene, encoding for a transcription factor HNF1α, which is a master regulator of plasma protein fucosylation. The presence of loss-of-function variants in the *HNF1A* gene leads to upregulation of core fucosylation and downregulation of antennary fucosylation of N-glycans. A new liquid chromatography with tandem mass spectrometry (LC-MS/MS) method was developed to assess fucosylation levels for 320 patient blood plasma samples and evaluate the biomarker's diagnostic performance across different research centres. The results showed a strong correlation of the measured fucosylation levels between the two centres, with correlation coefficients of up to 0.88 for the relevant glycosylation traits. The improved chromatographic separation allowed for the identification of six single glycan traits and a derived antennary fucosylation trait, which effectively differentiated between pathogenic mutation carriers and those with benign or no mutations,

achieving an area under the curve (AUC) of up to 0.94 in the receiver operating characteristic curve analysis.

The subsequent study introduces an enzymatic plate-based assay for assessing α1-3,4 fucosylation levels, aiming to overcome the applicability bottleneck of LC methods in public diagnostic centres (**Chapter 3**). Previous research has demonstrated the effectiveness of the *N*-glycan biomarker in differentiating HNF1A-MODY cases using LC methods. In the current study, a high-throughput exoglycosidase plate-based assay was developed to measure α1-3,4 fucosylation levels in blood plasma samples. This assay was optimised and validated with 1000 clinical samples from a cohort of young-adult onset diabetes patients, including HNF1A-MODY and type 2 diabetes cases. The α1-3,4 fucosylation levels effectively differentiated cases with pathogenic *HNF1A* variants, achieving an AUC value of 0.87. This method was evaluated against previously applied LC methods.

In the second research part, altered apolipoprotein CIII (apo-CIII) *O*-glycosylation profiles linked to increased plasma triglyceride levels in diabetic dyslipidemia are investigated. A highly-automated ultra-high resolution MALDI-FTICR MS method for analysing intact apo-CIII was optimized for large sample cohorts incorporating a chemical oxidation step to reduce methionine oxidation heterogeneity and spectrum complexity (**Chapter 4**). Sinapinic acid matrix minimised sialic acid loss during MALDI measurements, and the MassyTools software was applied for standardised and automated MS data processing and quality control. When applied to 771 plasma samples from individuals without diabetes to access relative levels of apo-CIII glycoforms, the method validated the relationship between apo-CIII glycoforms and lipid biomarkers. The study supports the hypothesis that apo-CIII sialylation influences triglyceride clearance, independent of body mass index.

Subsequently, the proposed MALDI MS-based analytical workflow was applied to a large patient cohort involving 2318 participants from the DiaGene study, including participants with and without type 2 diabetes (**Chapter 5**). This study aimed to unravel

how apo-CIII glycosylation impacts lipid traits and associates with type 2 diabetes prevalence, and to explore the genetic basis of these effects through a genome-wide association study (GWAS). Variants in the *GALNT2* gene, linked to the overexpression of *O*-glycosylating enzyme GALNT2, and the *IFT172* gene were associated with specific apo-CIII glycosylation patterns, high-density lipoprotein (HDL) cholesterol and triglyceride levels. High non-glycosylated apo-CIII (apo-$CIII_{0a}$) levels were associated with increased HDL cholesterol and triglycerides, whereas, disialylated apo-CIII (apo-$CIII_2$) itself was linked to lower triglyceride levels. Replication of these genetic associations with lipid levels was performed in an additional cohort of 5409 individuals from the Diabetes Care System.

The final study investigates associations between apo-CIII glycosylation, genetic variants, and micro- and macrovascular complications of diabetes (retinopathy, nephropathy, neuropathy, cardiovascular disease) in two cohorts: the DiaGene study, n = 1571 and the Hoorn DCS cohort, n = 5409 (**Chapter 6**). Mono-sialylated apo-CIII (apo-$CIII_1$) and disialylated apo-CIII (apo-$CIII_2$) were associated with a reduced and increased risk of retinopathy, respectively. The *GALNT2* gene variant rs4846913, which is associated with lower glycosylated apo-CIII (apo-$CIII_{0a}$) levels, was linked to a decreased prevalence of retinopathy. Higher levels of apo-$CIII_1$ were associated with an increased risk of neuropathy and lower levels of apo-$CIII_{0a}$ were associated with a reduced risk of macrovascular complications.

This thesis concludes with critically evaluating the research findings and their potential clinical implications. Challenges in translating glycan biomarkers into clinical practice are discussed, along with the evaluation of analytical methods and statistical approaches for biopharma industry use.

Nederlandse Samenvatting

In het onderzoek gepresenteerd in dit proefschrift ligt de nadruk op het onderzoek naar biomarkers in diabetes, dat wil zeggen de ziekte-gerelateerde parameters, de mogelijke klinische implicaties hiervan en de ontwikkeling van analytische methodes. Glycan biomarkers zijn veelbelovende hulpmiddelen voor de diagnose, prognose en het monitoren van verschillende ziekten, waaronder kanker, auto-immuunziekten en infectieziekten. In dit proefschrift ligt de focus op de ontwikkeling en optimalisering van analytische methoden voor het onderzoeken van veranderingen in glycosylerings profielen als potentiële biomarkers in grootschalige klinische cohorten van patiënten met diabetes en gerelateerde complicaties van deze ziekte.

Het doel van het onderzoek richt zich op het aanpakken van de uitdagingen die zich uiten tijdens high-throughput monsterbereiding voor de analyse van glycosylering, data verwerking en statistische analyse. De ontwikkelde methoden zijn toepasbaar voor het analyseren van *N*-glycanen, *O*-glycosyleerde eiwitten en voor de bepaling van absolute fucosylatieniveaus van eiwitten in bloedplasma.

Het proefschrift leidt in met een algemene beschrijving over de verschillende soorten glycosylering, glyco(proteo)mics analytische benaderingen en de ontdekking van glycan biomarkers in diabetes (**Hoofdstuk 1**). Het onderzoek begint met een studie naar de *N*-glycan antenne fucosylering als biomarker voor de monogene vorm van diabetes, HNF1A-MODY (**Hoofdstuk 2**). Individuen met HNF1A-MODY zijn dragers van varianten in het *HNF1A* gen, dit gen codeert de transcriptie factor HNF1α, welk een cruciaal onderdeel is voor de regulatie van plasma eiwit fucosylatie. De aanwezigheid van de varianten in het *HNF1A* gen leidt tot verhoging van core fucosylatie and verlaging van antenne fucosylatie van *N*-glycanen. De ontwikkeling van een nieuwe vloeistofchromatografie in combinatie met tandemmassaspectrometrie (LC-MS/MS) methode voor de bepaling van fucosyleringsniveaus in 320 bloedplasmasamples van diabetes patiënten en om de diagnostische potentie van de biomarker te evalueren

tussen twee verschillende onderzoekscentra. De resultaten lieten een sterke correlatie van de gemeten fucosyleringsniveaus zien tussen de twee centra, met correlatiecoëfficiënten tot 0,88 voor de relevante glycosylerings profielen. De verbeterde chromatografische scheiding maakte identificatie van zes glycaan kenmerken en een afgeleide antenne fucosyleringskenmerk mogelijk. De bevindingen waren effectief in het differentiëren van dragers van pathogene mutaties en degenen met goedaardige of geen mutaties, met een '*area under the curve*' (AUC) tot 0,94 in een signaaldetectietheorie analyse (*receiver operating characteristic* (ROC) curve-analyse).

De daaropvolgende studie introduceert een enzymatische test op plaat voor het beoordelen van α1-3,4 fucosyleringsniveaus, met als doel de toepasbaarheid obstakels van LC-methodes in een klinische setting te overwinnen (**Hoofdstuk 3**). Vorig onderzoek heeft de effectiviteit van de *N*-glycan biomarkers aangetoond bij het differentiëren van HNF1A-MODY patiënten met behulp van LC methoden. In de huidige studie werd een high-throughput exoglycosidase op een plaat test ontwikkeld om α1-3,4 fucosyleringsniveaus in bloedplasmasamples te meten. De methode werd geoptimaliseerd en gevalideerd met een cohort van 1000 klinische monsters van jong-volwassen patiënten met diabetes , inclusief HNF1A-MODY en type 2 diabetes samples. De α1-3,4 fucosyleringsniveaus differentieerden effectief gevallen met pathogene *HNF1A*-varianten, met een AUC-waarde van 0,87. Deze methode werd geëvalueerd ten opzichte van eerder beschreven LC-methoden.

In het tweede deel van het onderzoek worden veranderde apolipoproteïne CIII (apo-CIII) *O*-glycaan profielen die verband houden met verhoogde triglyceride niveaus in plasma bij diabetische dyslipidemie onderzocht. Een gevorderde geautomatiseerde ultra-hoge resolutie MALDI-FTICR MS-methode voor het analyseren van intact apo-CIII werd geoptimaliseerd voor de analyse van grootschalige cohort geïntegreerd met een chemische oxidatiestap om methionine-oxidatie heterogeniteit en spectrumcomplexiteit te verminderen (**Hoofdstuk 4**). Sinapinic acid matrix

minimaliseerde het verlies van siaalzuur tijdens de MALDI-meting, en de software, MassyTools werd toegepast voor gestandaardiseerde en geautomatiseerde MS-dataverwerking en kwaliteitscontrole. De methode werd toegepast op 771 plasmasamples van individuen zonder diabetes om relatieve niveaus van apo-CIII glycovormen te beoordelen, dit valideerde de methode met betrekking tot de relatie tussen apo-CIII glycovormen en lipidenbiomarkers. De studie ondersteunt de hypothese dat apo-CIII sialylering triglyceride opruiming beïnvloedt, onafhankelijk van de body mass index.

Vervolgens werd de geoptimaliseerde MALDI MS-gebaseerde analytische workflow toegepast op een grootschalige patiënten cohort met 2318 deelnemers uit de DiaGene studie, zowel met als zonder type 2 diabetes (**Hoofdstuk 5**). Deze studie had als doel te ontrafelen hoe apo-CIII glycosylering lipidenkenmerken beïnvloedt en associeert met de prevalentie van type 2 diabetes, en om de genetische basis van deze effecten te verkennen via een genoom associatiestudie (genome wide association study, GWAS). Varianten in het *GALNT2*-gen, die verband worden gebracht met de over expressie van het *O*-glycosylerende enzym GALNT2, en het *IFT172*-gen werden geassocieerd met specifieke apo-CIII glycosylerings patronen, high-density-lipoproteïne (HDL) cholesterol en triglycerideniveaus. Hoge niveaus van niet-glycosyleerde apo-CIII (apo-$CIII_{0a}$) werden geassocieerd met verhoogde HDL-cholesterol en triglyceriden, terwijl disialylering van apo-CIII (apo-$CIII_2$) zelf werd gekoppeld aan lagere triglycerideniveaus. Replicatie van deze genetische associaties met de lipiden niveaus werd uitgevoerd in een extra cohort van 5409 individuen uit het Diabetes Care System.

In het laatste gedeelte van het proefschrift wordt een studie beschreven met associaties tussen apo-CIII glycosylering, genetische varianten, en micro- en macrovasculaire complicaties van diabetes (retinopathie, nefropathie, neuropathie, cardiovasculaire ziekte) in twee cohorten: de DiaGene studie, n = 1571 en de Hoorn DCS cohort, n = 5409 (**Hoofdstuk 6**). Mono-sialylering apo-CIII (apo-$CIII_1$) en disialylering apo-CIII (apo-$CIII_2$) werden geassocieerd met respectievelijk een verminderd en

verhoogd risico op retinopathie. De *GALNT2*-genvariant rs4846913, die geassocieerd is met lagere glycosyleerde apo-CIII (apo-$CIII_{0a}$) niveaus, werd gekoppeld aan een verminderd prevalentie van retinopathie. Hogere niveaus van apo-$CIII_1$ werden geassocieerd met een verhoogd risico op neuropathie en lagere niveaus van apo-$CIII_{0a}$ werden geassocieerd met een verminderd risico op macrovasculaire complicaties.

Het proefschrift wordt afgesloten met een kritische evaluatie van de onderzoeksresultaten en hun potentiële klinische implicaties. Uitdagingen bij het vertalen van glycan biomarkers naar de klinische praktijk worden besproken, samen met de evaluatie van analytische methoden en statistische benaderingen voor gebruik in de biofarmaceutische industrie.

Curriculum vitae

Daniel A. Demus was born on the 2nd of February 1991 in Lubaczów, Poland. He pursued his passion for biomedical science during an undergraduate degree program, earning a Bachelor of Engineering in Biotechnology with a specialization in biomaterials and biophysical approaches for nanostructure research at Rzeszów University in Poland in 2014. Subsequently, he continued his education by enrolling in the Master of Science in Biotechnology program at the University of Life Sciences in Lublin, Poland.

Daniel gained his initial industrial experience in a GMP-regulated department at Merck, Germany, in 2014, during a one-year internship during his Master's degree. This experience provided him with a solid understanding of quality control and processes within a large organization. He successfully completed his Master of Science in Biotechnology degree in June 2016, with his master's thesis focusing on whole cell proteome profiling of bacterial species isolated from poultry, utilizing MALDI-TOF mass spectrometry. This research was part of collaborative efforts between the Biotechnology and Veterinary faculties of the University of Life Sciences in Lublin and external partners, aiming to advance the development of animal vaccines.

With a growing interest in the application of mass spectrometry in biomedical science, Daniel joined Ludger in Oxford, UK, in February 2017, as part of the European Industrial Doctorate training program "GlySign" (www.glysign.eu). This program was initiated in collaboration with Leiden University Medical Centre (LUMC), Netherlands, and Genos, Croatia. He conducted research projects studying glycan-based stratification biomarkers in diabetes mellitus and was involved in other GlySign network projects addressing cancer and autoimmune diseases. He participated in annual network meetings and completed a 12-month secondment at LUMC between 2018 and 2019. The highly collaborative nature of the network, bridging academia and industry, provided him with extensive experience in the development and optimization of analytical methods for glycosylation and glycoproteomics analysis in large clinical

sample cohorts, with a focus on method robustness, data quality, and compliance with good clinical practice.

In March 2022, Daniel joined a large molecule characterization team as a senior scientist at Lonza in the UK, gaining an opportunity to explore the biologics field in a manufacturing organization. His next career move within the industry was taking on the role of an analytical lead for biologics at GSK in the UK.

Daniel holds major interests in translational research, precision medicine, biomedical innovations, biologics discovery, and evidence-based wellness.

List of publications

1. **Demus D**, Jansen BC, Gardner RA, Urbanowicz PA, Wu H, Štambuk T, Juszczak A, Medvidović EP, Juge N, Gornik O, Owen KR, Spencer DIR. *Interlaboratory evaluation of plasma N-glycan antennary fucosylation as a clinical biomarker for HNF1A-MODY using liquid chromatography methods*. **Glycoconj J. 2021** Jun;38(3):375-386. doi: 10.1007/s10719-021-09992-w. Epub 2021 Mar 25. PMID: 33765222; PMCID: PMC8116301.
2. **Demus D**, Urbanowicz PA, Gardner RA, Wu H, Juszczak A, Štambuk T, Medvidović EP, Owen KR, Gornik O, Juge N, Spencer DIR. *Development of an exoglycosidase plate-based assay for detecting α1-3,4 fucosylation biomarker in individuals with HNF1A-MODY.* **Glycobiology. 2022** Mar 30;32(3):230-238. doi: 10.1093/glycob/cwab107. PMID: 34939081; PMCID: PMC8966479.
3. **Demus D**, Naber A, Dotz V, Jansen BC, Bladergroen MR, Nouta J, Sijbrands EJG, Van Hoek M, Nicolardi S, Wuhrer M. *Large-Scale Analysis of Apolipoprotein CIII Glycosylation by Ultrahigh Resolution Mass Spectrometry.* **Front Chem. 2021** May 7;9:678883. doi: 10.3389/fchem.2021.678883. PMID: 34026735; PMCID: PMC8138127.
4. Naber A, **Demus D**, Slieker R, Nicolardi S, Beulens JWJ, Elders PJM, Lieverse AG, Sijbrands EJG, 't Hart LM, Wuhrer M, van Hoek M. *Apolipoprotein-CIII O-Glycosylation, a Link between GALNT2 and Plasma Lipids.* **Int J Mol Sci. 2023** Oct 2;24(19):14844. doi: 10.3390/ijms241914844. PMID: 37834292; PMCID: PMC10573541.
5. Naber A, **Demus D**, Slieker RC, Nicolardi S, Beulens JWJ, Elders PJM, Lieverse AG, Sijbrands EJG, 't Hart LM, Wuhrer M, van Hoek M. *Apolipoprotein-CIII O-Glycosylation Is Associated with Micro- and Macrovascular Complications of Type 2 Diabetes.* **Int J Mol Sci. 2024** May 14;25(10):5365. doi: 10.3390/ijms25105365. PMID: 38791405; PMCID: PMC11121677.

Acknowledgments

It has been a journey with highs and lows, filled with many amazing experiences, memories and people.

First of all, I would like to thank my supervisors - my promotor and co-promotors. Prof. Manfred Wuhrer, thank you for your scientific guidance, commitment, and enthusiasm. Your clear vision and ability to always find solutions are inspiring! Dr. Daniel Spencer, thank you for guiding me to build a strong foundation and develop solid analytical skills. Your kindness and understanding throughout and after GlySign have been priceless. Dr. Mandy van Hoek, thank you for bringing the clinical perspective, showing us your dedication to patients, and creating great collaborations, bringing even more talented scientists and clinicians to our studies.

To the entire Ludger crew, it was truly the best time being part of a team full of diverse and colourful personalities, learning from each other every day. To the CPM crew, the best brains in the glyco-field, I am proud to be part of such an exciting, hardworking research group with one of the strongest analytical capabilities I have seen and, surely, many more successes coming in the future. To LUMC, the clinical-academic spirit of this place is truly impressive. There are good and bad days in the office, but seeing patients in the buildings every day made me approach my work differently. I am so thankful to have experienced that.

To all my daily supervisors, every day is a new chapter, and we have created many such chapters to revise and learn from. Each discussion, agreement, and disagreement has pushed things to the next level. You have done a great job, and your experience and curiosity have been inspiring. Helen and Manu, our best teachers, thank you for bringing even more sunshine to GlySign, and for reminding us about the bright side of the scientific and non-scientific world.

To the GlySigners (early-stage researchers at the time, now professional scientists), thank you for making it so beautiful and special. Seeing a real potential in the research, assays, and outreach activities we worked on was a drive. Everyone was so focused, pushing each other in our unique ways as we all wanted to succeed individually and together for the patients.

To old and new friends I have made along the way, Iwona and Aniela, it's difficult to describe, but thank you for everything, PhD and non-PhD related. And the friendship lasts!

To my family, my parents, thank you for trusting my choices from day one, letting me choose my ways and simply be myself. To my brother for always having my back.

To all scientists, clinicians, patients, and volunteers involved in the studies presented in this PhD thesis, thank you for making this happen and contributing another building block to science. To everyone involved in preparing and evaluating this thesis, and my wonderfully helpful Paranymphs, we all own it proudly. Let's continue making science, keep pushing every day to see more, understand better, and evaluate more fully.

A learning and growing journey. Thank you!