Universiteit
Leiden
The Netherlands

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

Demus, D.A.

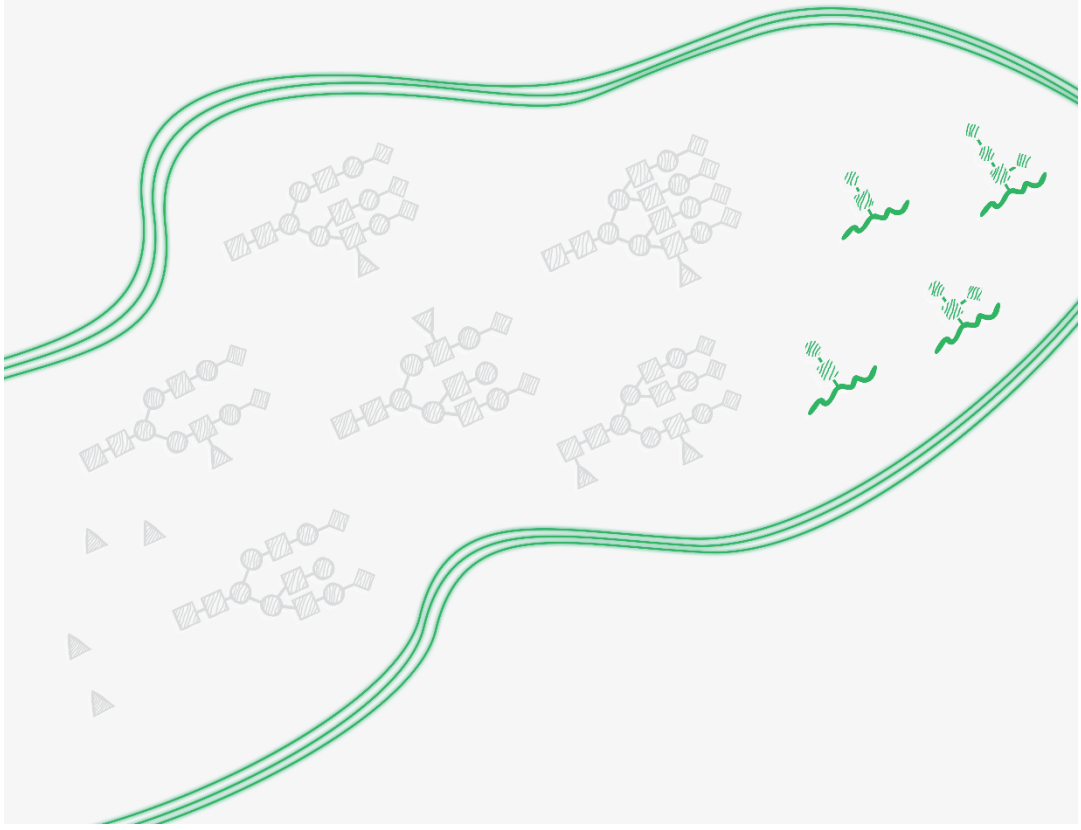
Citation

Demus, D. A. (2024, October 1). *Biomarker discovery in diabetes mellitus and lipid metabolism: multi-platform glyco(proteo)mic approaches*. Retrieved from https://hdl.handle.net/1887/4093481

Version:	Publisher's Version
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/4093481

Note: To cite this publication please use the final published version (if applicable).

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

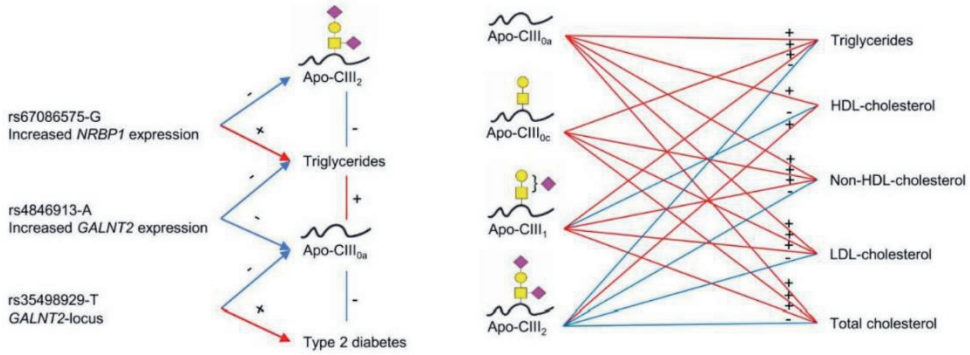
Reprinted and adapted: Naber, A.; Demus, D.; Slieker, R.; Nicolardi, S.; Beulens, J.W.J.; Elders, P.J.M.; Lieverse, A.G.; Sijbrands, E.J.G.; 't Hart, L.M.; Wuhrer, M.; et al. Apolipoprotein-CIII O-Glycosylation, a Link between GALNT2 and Plasma Lipids. Int. J. Mol. Sci. 2023, 24, 14844. https://doi.org/10.3390/ijms241914844

Copyright © 2023 The Authors

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



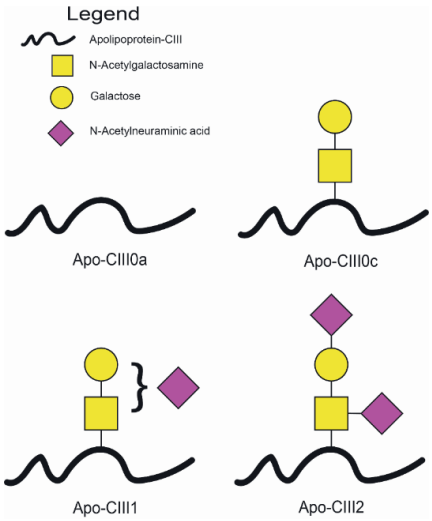


Figure S1. The four most abundant apo-CIII proteoforms. For graphical representations of glycan structures: yellow square (*N*-acetylgalactosamine), yellow circle (galactose), purple diamond (*N*-acetylneuraminic acid).

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-CII0a	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-CII0c	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-CII1	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-CII2	-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-CII1/apo-CII2 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-CII0a	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-CII0c	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-CII1	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-CII2	-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-CII1/apo-CII2 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**).

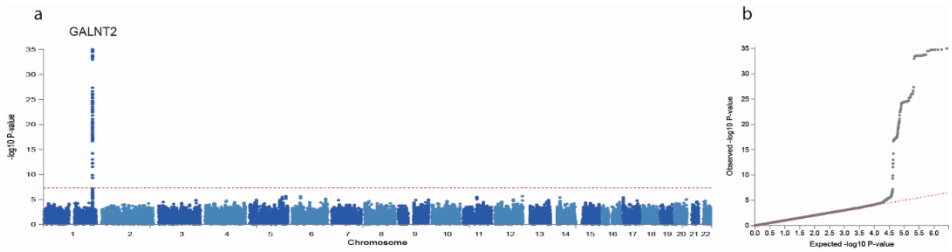


Figure 1. Manhattan and QQ plot of apo-CIII$_{0a}$ GWAS. GWAS of apo-CIII$_{0a}$ showing associations with alleles at the *GALNT2* locus on chromosome 1. Manhattan plot (left) showing significance of the association of each SNP allele with apo-CIII$_{0a}$ by plotting the -log10 of the p-value against 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	0.002928	7.07×10^{-7}	*PPFIBP1* (intronic)
Apo-$CIII_2$	rs67086575	2	27686480	G	A	0.1502	−0.0469	0.009473	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	0.006647	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	0.006713	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	0.008028	1.80×10^{-7}	*RN7SKP181* (intergenic)
	rs10412211	19	13591542	T	G	0.4579	0.0396	0.008031	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.

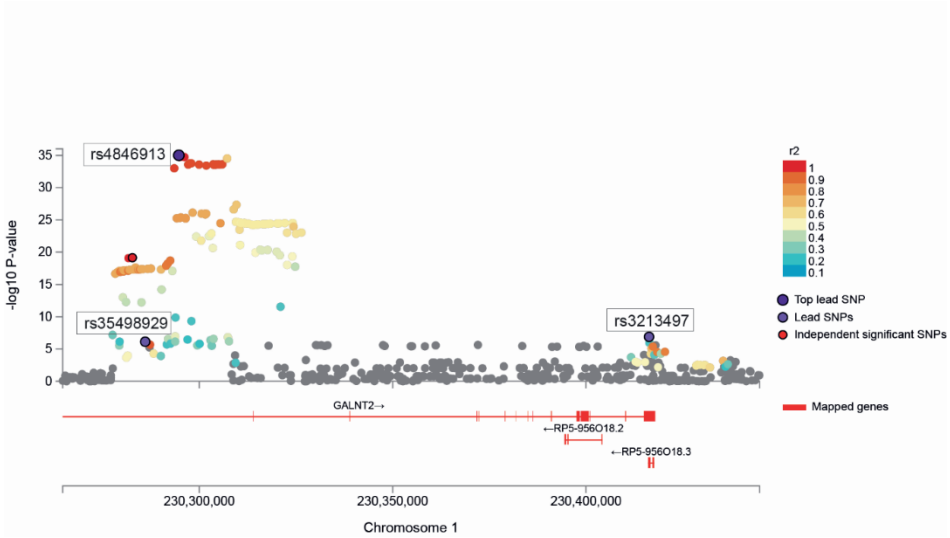


Figure S3. Region plot for the A allele of rs4846913 at the *GALNT2* locus on chromosome 1.

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**).

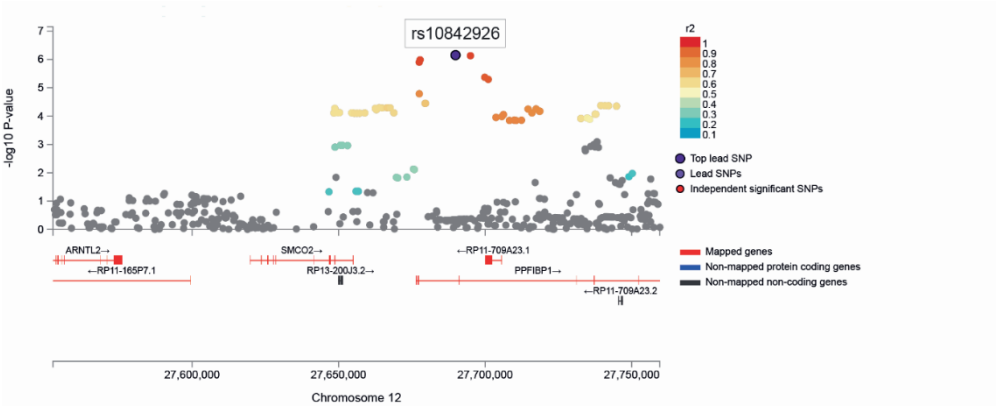


Figure S4. Region plot for the rs10842926-C in the *PPFIBP1* locus.

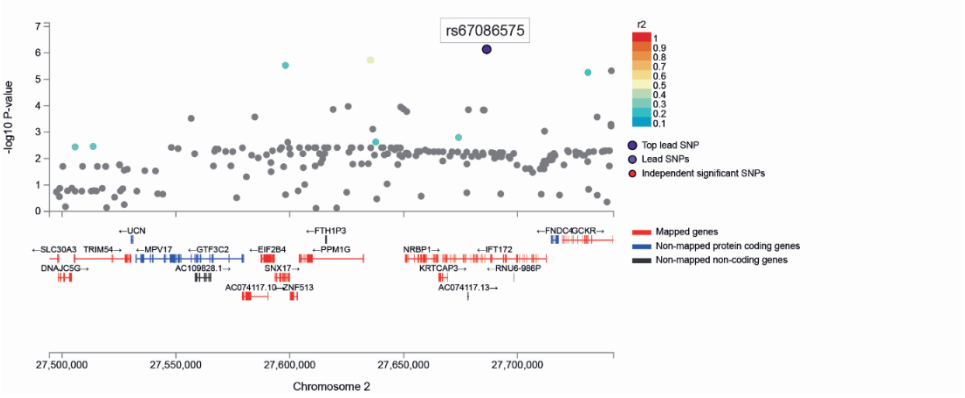


Figure S5. Region plot for the rs67086575-G allele in the intronic region of the IFT172-gene.

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.