Human Molecular Genetics, 2011, Vol. 20, No. 24 **5000–5011**
doi:10.1093/hmg/ddr414
Advance Access published on September 9, 2011

Polymorphisms in *B3GAT1*, *SLC9A9* and *MGAT5* are associated with variation within the human plasma *N*-glycome of 3533 European adults

Jennifer E. Huffman[1,*], Ana Knežević[2], Veronique Vitart[1], Jayesh Kattla[3], Barbara Adamczyk[3], Mislav Novokmet[2], Wilmar Igl[4], Maja Pučić[2], Lina Zgaga[5], Åsa Johannson[4], Irma Redžić[6], Olga Gornik[6], Tatijana Zemunik[7], Ozren Polašek[7], Ivana Kolčić[7], Marina Pehlić[7], Carolien A.M. Koeleman[8], Susan Campbell[1], Sarah H. Wild[5], Nicholas D. Hastie[1], Harry Campbell[5], Ulf Gyllensten[4], Manfred Wuhrer[8], James F. Wilson[5], Caroline Hayward[1], Igor Rudan[5,7], Pauline M. Rudd[3], Alan F. Wright[1] and Gordan Lauc[2,6,*]

[1]MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, Western General Hospital, Edinburgh EH4 2XU, UK, [2]Glycobiology Division, Genos Ltd, Planinska 1, Zagreb 10000, Croatia, [3]Dublin–Oxford Glycobiology Laboratory, National Institute for Bioprocessing Research and Training, Conway Institute, University College Dublin, Belfield, Dublin 4, Ireland, [4]Department of Genetics and Pathology, Rudbeck Laboratory, Uppsala University, Uppsala 751 85, Sweden, [5]Centre for Population Health Sciences, The University of Edinburgh Medical School, Edinburgh, UK, [6]Faculty of Pharmacy and Biochemistry, University of Zagreb, Ante Kovačića 1, Zagreb 10000, Croatia, [7]Faculty of Medicine, University of Split, Soltanska 2, Split 21000, Croatia and [8]Biomolecular Mass Spectrometry Unit, Department of Parasitology, Leiden University Medical Center, Leiden, The Netherlands

Received June 29, 2011; Revised and Accepted September 7, 2011

The majority of human proteins are post-translationally modified by covalent addition of one or more complex oligosaccharides (glycans). Alterations in glycosylation processing are associated with numerous diseases and glycans are attracting increasing attention both as disease biomarkers and as targets for novel therapeutic approaches. Using a recently developed high-throughput high-performance liquid chromatography (HPLC) analysis method, we have reported, in a pilot genome-wide association study of 13 glycan features in 2705 individuals from three European populations, that polymorphisms at three loci (*FUT8*, *FUT6*/*FUT3* and *HNF1A*) affect plasma levels of *N*-glycans. Here, we extended the analysis to 33 directly measured and 13 derived glycosylation traits in 3533 individuals and identified three novel gene association (*MGAT5*, *B3GAT1* and *SLC9A9*) as well as replicated the previous findings using an additional European cohort. *MGAT5* (meta-analysis association *P*-value = 1.80×10^{-10} for rs1257220) encodes a glycosyltransferase which is known to synthesize the associated glycans. In contrast, neither *B3GAT1* (rs7928758, $P = 1.66 \times 10^{-08}$) nor *SLC9A9* (rs4839604, $P = 3.50 \times 10^{-13}$) had previously been associated functionally with glycosylation of plasma proteins. Given the glucuronyl transferase activity of B3GAT1, we were able to show that glucuronic acid is present on antennae of plasma glycoproteins underlying the corresponding HPLC peak. *SLC9A9* encodes a proton pump which affects pH in the endosomal compartment and it was recently reported that changes in Golgi pH can impair protein sialylation, giving a possible mechanism for the observed association.

*To whom correspondence should be addressed. Tel: +44 1313322471; Fax: +44 1314678456; Email: jennifer.huffman@hgu.mrc.ac.uk (J.E.H.); Fax: +385 16394400; Email: glauc@pharma.hr (G.L.)

INTRODUCTION

Glycosylation is a post-translational modification that influences protein and lipid structural complexity and function. It is essential for multicellular life and its complete absence is embryonic lethal (1). Genetic defects that affect glycosylation biosynthetic or degradation pathways have been associated with a wide range of congenital disorders (2,3). Multiple proteins are modified by covalently bound glycans, which has implications for protein folding, degradation, cell signalling, secretion, immune function and transcription (4,5). Misregulation of glycosylation has been associated with a wide spectrum of common diseases including cancer, diabetes, cardiovascular, immunological and infectious disorders (4,6,7). Terminal variability in glycans is common (e.g. ABO blood groups) and helps to create diversity that allows our bodies to evade pathogens and adapt to changing environments (8). Glycosylation is not directly template-driven, which gives rise to the extensive complexity of the glycoproteome, estimated to be several orders of magnitude more complex than the proteome itself (9).

Owing to experimental limitations in quantifying glycans in biological samples, our understanding of the genetic regulation of glycosylation is still very limited (2). Recent technological advances have allowed reliable, high-throughput quantification of *N*-glycans (10), which now allows investigation into the genetic regulation and biological role of these glycan structures and brings glycomics into line with genomics, proteomics and metabolomics (11). In contrast to other analytical techniques, high-performance liquid chromatography (HPLC) separates glycans into chromatographic peaks which contain mostly similar structures (Fig. 1). Therefore, even without complete separation of all glycan structures, many glycome features of an individual sample can be deduced from ratios of different HPLC peaks (12).

An initial pilot study was performed using three cross-sectional population samples [CROATIA-Vis, CROATIA-Korcula and ORCADES (The Orkney Complex Disease Study)] on a subset of directly measured [desialylated glycans 1–13 (DG1–DG13)] and derived glycan species [antennary fucosylated glycans (FUC-A), core fucosylated glycans (FUC-C)] to investigate whether a genome-wide association study (GWAS) approach could be applied to the identification of genetic variants affecting *N*-glycans in human plasma. Single-nucleotide polymorphisms (SNPs) located within *fucosyltransferase 8* (*FUT8*), *fucosyltransferase 6* (*FUT6*) and *hepatic nuclear factor 1 alpha* (*HNF1a*) genes were found to be significantly associated with levels of several glycan groups (13).

In this study, a similar analysis was performed to include additional glycan species and an additional population sample (the Northern Swedish Population Health Study: NSPHS). We were able to strengthen our previous results as well as describe additional associations of these genes with new glycan traits and novel associations with *mannosyl* (*α-1,6-*)*-glycoprotein β-1,6-N-acetyl-glucosaminyltransferase V* (*MGAT5*), *β-1,3-glucuronyltransferase 1* (*B3GAT1*), and *solute carrier family 9, member 9* (*SLC9A9*).

RESULTS

A meta-analysis of genome-wide association results was conducted for 46 plasma *N*-glycan traits (Table 1) measured in four population-based samples, CROATIA-Vis, CROATIA-Korcula, ORCADES and NSPHS. Sex-specific analyses were also performed for each trait, and descriptive statistics and heritabilities for all traits are presented in Supplementary Material, Table S1. Heritability of traits (adjusted for sex and age) was variable between populations, as might be expected given the isolated nature of the populations. However, all traits had at least one population displaying a heritability >0.2, with most having at least one population in which heritability was >0.4. DG10 showed a fairly consistent heritability across populations ($h^2 = 0.31–0.40$), whereas GP10 was much more variable ($h^2 = 0.21–0.65$). Meta-analysis summary data for each significantly associated gene region are presented in Table 1. Summary data for all genome-wide significant SNPs are presented in Supplementary Material, Tables S2 (all participants), S3 (women only) and S4 (men only).

The most statistically significant association was found with a pair of SNPs located on chromosome 19. rs3760776 and rs778805 were both significantly associated with DG7, DG9, GP14, and only rs3760776 reached genome-wide significance with DG12 and A-FUC. The association of rs3760776 with DG7, DG9, DG12 and A-FUC was reported in the previous study, and the addition of another population sample improved the association *P*-value for each trait (13). The most significant association was found with DG9 for SNP rs3760776 ($P = 3.2 \times 10^{-29}$), located at the 5′ end of the *fucosyltranferase 6* (*FUT6*, Entrez GeneID: 2528). The effect size of the G allele of rs3760776 was the largest of all traits, 0.44 (SE 0.04) for DG9 [*z*-score units, after adjustment for sex, age and principal components (PC)]. The association region for this SNP contains the *NRTN*, *FUT6* and *FUT3* genes between two recombination hotspots (Supplementary Material, Fig. S1). *FUT6* encodes the enzyme fucosyltransferase VI, which was reported to be the key enzyme responsible for the antennary fucosylation of plasma proteins (14), thus the causal variant(s) probably affect this gene. SNP rs3760776 explained 1.1, 4.8, 5.7 and 3.0% of the variance of DG9 (adjusted for sex, age and PC) in CROATIA-Vis, CROATIA-Korcula, ORCADES and NSPHS, respectively.

An association peak containing 22 SNPs was located in the region encompassing the *FUT8* (Entrez GeneID: 2530) gene on chromosome 14 (Supplementary Material, Fig. S2). We previously reported associations of this gene with plasma concentrations of DG1 and DG6. The most significant signal for DG1 was still located 5′ of the gene with the addition of the NSPHS population sample; however, the SNP showing the strongest association changed from rs7159888 in the previous study to rs11621121 whose *P*-value decreased from 1.0×10^{-17} to 1.7×10^{-23}. The association with DG6 produced the same top SNP (rs10483776), with its *P*-value decreasing from 9.6×10^{-09} to 9.3×10^{-10}. SNPs in this region were also associated with DG10, GP1, GP10, C-FUC and A2 in a manner consistent with the biological role of FUT8. The SNP rs11621121 explained 3.0, 5.9, 1.0 and 2.8% of the variance of DG1

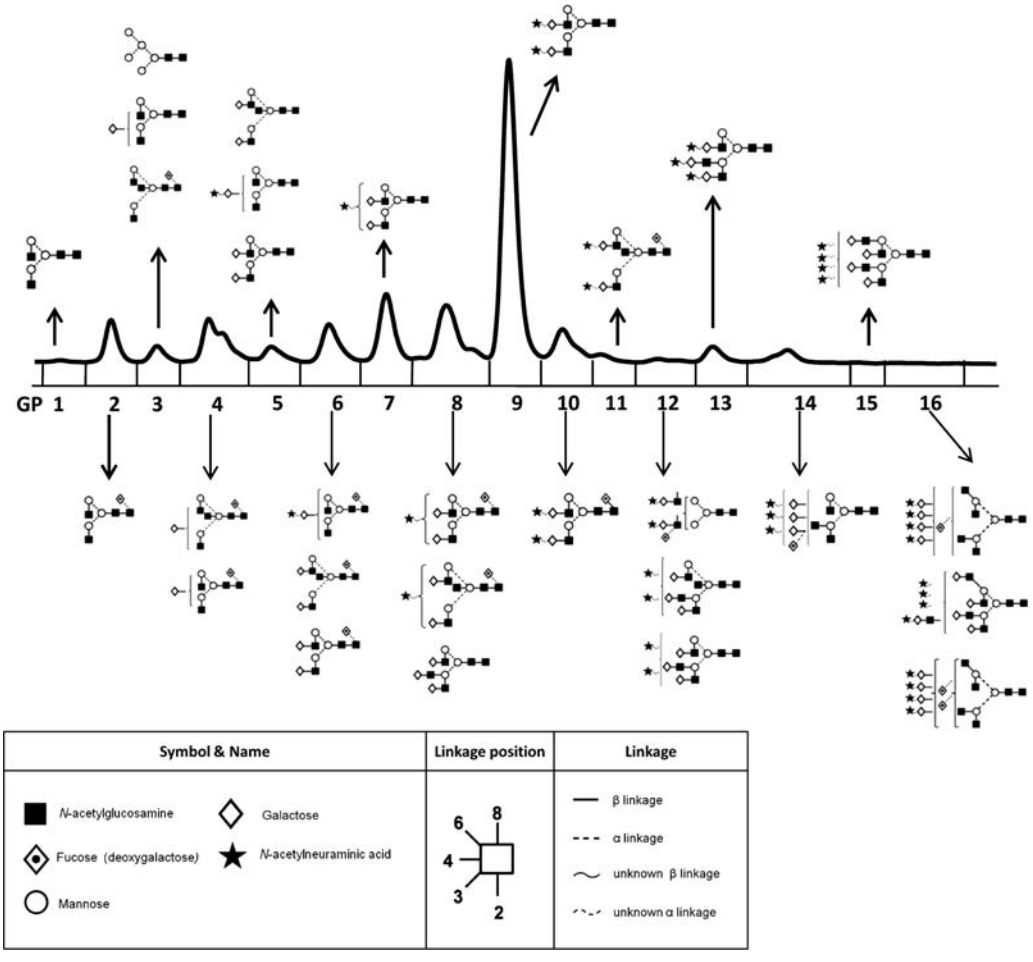

Figure 1. Structures of glycans separated by HPLC-HILIC analysis of the plasma glycome. Structures are given using Oxford notation (62).

(adjusted for sex, age and PC) in CROATIA-Vis, CROATIA-Korcula, ORCADES and NSPHS, respectively.

Five SNPs on chromosome 12, the most significant being rs735396, showed genome-wide significant associations with DG7 ($P = 7.8 \times 10^{-12}$). SNPs within this region were also associated with GP13 ($P = 3.8 \times 10^{-10}$), GP15 ($P = 4.3 \times 10^{-08}$), DG9 ($P = 3.2 \times 10^{-11}$), DG11 ($P = 4.2 \times 10^{-09}$) and FUC-A ($P = 5.0 \times 10^{-10}$). Associations of DG7 and DG11 with SNPs in this region were previously reported but the significance of the association has increased with the addition of NSPHS, and the number of SNPs with significant association has also increased (13). All SNPs are located within or 5′ of the *HNF1 homeobox A* (*HNF1a*, Entrez GeneID: 6927) gene region, with rs735396 found in intron 9. Three other genes are located between the recombination hotspots that comprise the boundaries of the association interval, *C12orf27*, *C12orf43* and *OASL* (Supplementary Material, Fig. S3); however, functional studies reported in the previous paper have demonstrated that HNF1a plays a role in the regulation of several genes which play key roles in the fucosylation of glycan chains (13). SNP rs735396 explained 2.9, 0.9, 0.8 and 2.6% of the variance of sex-, age- and PC-adjusted DG7 in CROATIA-Vis, CROATIA-Korcula, ORCADES and NSPHS, respectively.

An SNP located on chromosome 2, rs1257220, was associated with DG11 and TA, a derived trait quantifying tetra-antennary glycans, ($P = 1.9 \times 10^{-10}$, $P = 1.8 \times 10^{-10}$). This SNP is located in the intron between exons 1 and 2 of the gene encoding mannosyl (α1,6)-glycoprotein β-1,6-*N*-acetyl-glucosaminyltransferase (*MGAT5*, Entrez GeneID: 4249). There is an area of high recombination in the middle of the *MGAT5*, indicating that the causal variant is most likely in the region 5′ of this hotspot (Fig. 2). The effect size of the A allele of rs1257220 is 0.19 (SE 0.03) for both DG11 and TA (*z*-score units, after adjustment for sex, age and PC). SNP rs1257220 explained 1.1, 1.6, 0.8 and 2.0% of the variance of sex-, age- and PC-adjusted TA in CROATIA-Vis, CROATIA-Korcula, ORCADES and NSPHS, respectively.

A single SNP located on chromosome 11, rs7928758, was associated with DG13 ($P = 1.7 \times 10^{-08}$). This SNP is located in the final intron of the *β-1,3-glucuronyltransferase1* (*B3GAT1*, Entrez GeneID: 27087*)* gene. There is a spike in the recombination rate towards the end of *B3GAT1*; therefore, the causal variant is most likely located 5′ of this spot (Fig. 3). The effect size of the C allele of rs7928758 is −0.23 (SE 0.04) (*z*-score units, after adjustment for sex, age and PC). SNP rs7928758 explained 0.2, 0.4, 2.4 and 1.6% of the variance of sex-, age- and PC-adjusted DG13 in CROATIA-Vis, CROATIA-Korcula, ORCADES and NSPHS, respectively. Given this association, we performed a mass spectrometric analysis of the 2AB-labeled glycans of three fractions of

Table 1. Genetic markers associated with plasma *N*-glycan levels

Trait	Chr	Lowest *P*-value SNP in associated interval	Meta-analysis *P*-value	Meta-analysis sample size	Associated interval size, kb (number of genome-wide significant SNPs in interval)	Genes within associated interval[a]	Major allele, minor allele (MAF)	Effect size for minor allele (SE)[b]
All participants								
GP1	14q23	rs7159888	2.06×10^{-21}	3242	650 (15)	*FUT8*	G, A (0.45)	0.25 (0.03)
GP10	14q23	rs10483776	1.03×10^{-14}	3269	650 (9)	*FUT8*	A, G (0.21)	−0.25 (0.03)
DG1	14q23	rs11621121	1.69×10^{-23}	3234	650 (22)	*FUT8*	A, G (0.43)	0.27 (0.03)
DG6	14q23	rs10483776	9.28×10^{-10}	3266	650 (6)	*FUT8*	A, G (0.21)	−0.20 (0.03)
DG10	14q23	rs10483776	1.50×10^{-08}	3261	650 (1)	*FUT8*	A, G (0.21)	−0.19 (0.03)
C-FUC	14q23	rs10483776	1.28×10^{-08}	3266	650 (1)	*FUT8*	A, G (0.21)	−0.19 (0.03)
A2	14q23	rs7159888	2.08×10^{-22}	3182	650 (18)	*FUT8*	G, A (0.45)	0.26 (0.03)
GP14	19p13	rs3760776	8.66×10^{-17}	3279	54 (2)	*NRTN; FUT6; FUT3*	G, A (0.13)	−0.33 (0.04)
DG7	19p13	rs3760776	1.34×10^{-21}	3236	54 (2)	*NRTN; FUT6; FUT3*	G, A (0.13)	−0.38 (0.04)
DG9	19p13	rs3760776	3.18×10^{-29}	3262	54 (1)	*NRTN; FUT6; FUT3*	G, A (0.13)	−0.44 (0.04)
DG12	19p13	rs3760776	4.56×10^{-19}	3260	54 (1)	*NRTN; FUT6; FUT3*	G, A (0.13)	−0.35 (0.04)
A-FUC	19p13	rs3760776	2.80×10^{-20}	3234	54 (1)	*NRTN; FUT6; FUT3*	G, A (0.13)	−0.36 (0.04)
GP13	12q24	rs735396	3.76×10^{-10}	3280	118 (3)	*HNF1A; C12orf43; OASL*	A, G (0.40)	0.17 (0.03)
GP15	12q24	rs735396	4.31×10^{-08}	3272	118 (1)	*HNF1A; C12orf43; OASL*	A, G (0.40)	0.15 (0.03)
DG7	12q24	rs735396	7.81×10^{-12}	3236	118 (5)	*HNF1A; C12orf43; OASL*	A, G (0.39)	−0.18 (0.03)
DG9	12q24	rs7953249	3.15×10^{-11}	3262	118 (5)	*HNF1A; C12orf43; OASL*	A, G (0.47)	0.17 (0.03)
DG11	12q24	rs735396	4.19×10^{-09}	3267	118 (1)	*HNF1A; C12orf43; OASL*	A, G (0.39)	0.16 (0.03)
A-FUC	12q24	rs735396	4.95×10^{-10}	3234	118 (5)	*HNF1A; C12orf43; OASL*	A, G (0.40)	−0.17 (0.03)
DG11	2q21	rs1257220	1.86×10^{-10}	3269	440 (1)	*MGAT5*	G, A (0.26)	0.19 (0.03)
TA	2q21	rs1257220	1.80×10^{-10}	3263	440 (1)	*MGAT5*	G, A (0.26)	0.19 (0.03)
DG13	11q25	rs7928758	1.66×10^{-08}	3233	102 (1)	*B3GAT1*	A, C (0.12)	−0.23 (0.04)
Tetrasialylated	3q24	rs4839604	3.50×10^{-13}	3320	125 (2)	*SLC9A9*	G, A (0.23)	0.22 (0.03)
Females only								
GP1	14q23	rs4073416	7.47×10^{-15}	1863	650 (13)	*FUT8*	A, G (0.43)	0.28 (0.03)
GP10	14q23	rs10483776	3.18×10^{-11}	1885	650 (3)	*FUT8*	A, G (0.21)	−0.28 (0.04)
DG1	14q23	rs11621121	1.72×10^{-16}	1876	650 (14)	*FUT8*	A, G (0.43)	0.28 (0.03)
A2	14q23	rs3742597	9.72×10^{-15}	1839	650 (14)	*FUT8*	A, G (0.28)	0.30 (0.04)
GP14	19p13	rs3760776	1.31×10^{-08}	1886	54 (1)	*NRTN; FUT6; FUT3*	G, A (0.22)	−0.29 (0.05)
DG7	19p13	rs3760776	7.80×10^{-10}	1875	54 (1)	*NRTN; FUT6; FUT3*	G, A (0.22)	−0.31 (0.05)
DG9	19p13	rs3760776	1.07×10^{-16}	1887	54 (1)	*NRTN; FUT6; FUT3*	G, A (0.22)	−0.43 (0.05)
DG12	19p13	rs3760776	2.39×10^{-13}	1888	54 (1)	*NRTN; FUT6; FUT3*	G, A (0.22)	−0.37 (0.05)
A-FUC	19p13	rs3760776	1.80×10^{-09}	1874	54 (1)	*NRTN; FUT6; FUT3*	G, A (0.22)	−0.30 (0.05)
DG11	2q21	rs1257220	3.46×10^{-09}	1892	440 (1)	*MGAT5*	G, A (0.27)	0.22 (0.04)
TA	2q21	rs1257220	2.65×10^{-09}	1890	440 (1)	*MGAT5*	G, A (0.26)	0.22 (0.04)
Males only								
GP1	14q23	rs7159888	9.25×10^{-11}	1370	650 (3)	*FUT8*	G, A (0.44)	0.26 (0.04)
DG1	14q23	rs6573604	1.76×10^{-12}	1360	650 (6)	*FUT8*	A, G (0.19)	0.35 (0.05)
A2	14q23	rs6573604	1.97×10^{-12}	1341	650 (5)	*FUT8*	A, G (0.19)	0.35 (0.05)
GP14	19p13	rs3760776	3.64×10^{-09}	1393	54 (1)	*NRTN; FUT6; FUT3*	G, A (0.14)	−0.34 (0.06)
DG7	19p13	rs3760776	4.35×10^{-13}	1361	54 (1)	*NRTN; FUT6; FUT3*	G, A (0.14)	−0.42 (0.06)
DG9	19p13	rs3760776	2.56×10^{-14}	1375	54 (1)	*NRTN; FUT6; FUT3*	G, A (0.14)	−0.44 (0.06)
A-FUC	19p13	rs3760776	3.31×10^{-12}	1360	54 (1)	*NRTN; FUT6; FUT3*	G, A (0.14)	−0.41 (0.06)
Tetrasialylated	3q24	rs1372288	3.43×10^{-08}	1404	125 (2)	*SLC9A9*	A, G (0.25)	0.25 (0.05)
GP1	14q23	rs4073416	7.47×10^{-15}	1863	650 (13)	*FUT8*	A, G (0.43)	0.28 (0.03)

[a]Results are reported for NCBI Build 36 and all SNPs are on '+' strand.
[b]*z*-Score units for trait adjusted for sex, age and PC.

DG13 and confirmed the presence of glucuronic acid on large, complex-type *N*-glycans (Supplementary Material, Fig. S4). Two examples of the obtained fragment spectra which revealed the diagnostic ion at mass-to-charge ratio (*m/z*) 542 corresponding to an *N*-glycan antenna decorated with glucuronic acid are shown in Supplementary Material, Figure S5.

A pair of SNPs on chromosome 3 were significantly associated with tetrasialylated glycans, rs1372288 ($P = 3.6 \times 10^{-13}$) and rs4839604 ($P = 3.5 \times 10^{-13}$). Both SNPs are found 3′ of the *solute carrier family 9*, *member 9* (*SLC9A9* Entrez GeneID: 285195). The effect size for the minor allele for both SNPs is 0.22 (SE 0.03) (*z*-score units, after adjustment for sex, age and PC). All significant SNPs fall between the 3′ end of the *SLC9A9* gene and just before the 3′ end of *CHST2* but not within *CHST2*, which is oriented tail-to-tail with *SLC9A9* (Fig. 4). There is a large spike in the recombination rate separating the region of association from the 3′ end of *CHST2* so the causal variant most likely affects falls in the region closer to *SLC9A9*. SNP rs4839604 explained 3.0, 1.5, 0.7 and 2.3% of the variance of tetrasialylated glycans

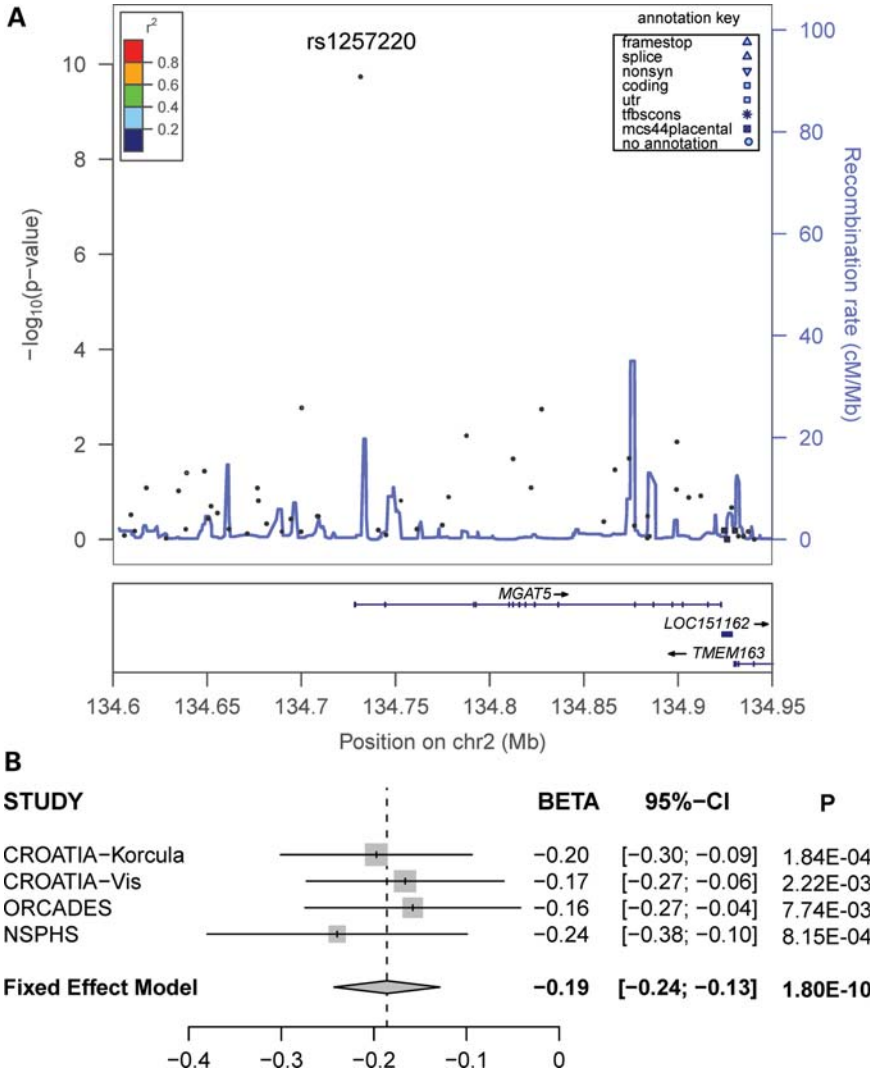

Figure 2. Significance plot (**A**) and forest plot (**B**) for the most significant SNP for regions of interest from the meta-analysis of tetra-antennary glycans (TA). (A) −Log 10 of the *P*-values are plotted against chromosome positions. The most significant SNP is labelled along with its *P*-value and information on SNP function (see annotation legend in the upper right corner). Gene information and regional LD (r^2) are also shown. (B) The estimate of the effect size of the best SNP for each individual population and the pool along with its standard error is shown, in which the size of the square for individual cohorts represents the proportion that it is contributing to the estimate of pooled effect size. The effect size presented is the β-coefficient, which represents a change in glycan levels measured in standard deviation units (adjusted for age, sex and PC) per copy of the allele modelled.

(adjusted for sex, age and PC) in CROATIA-Vis, CROATIA-Korcula, ORCADES and NSPHS, respectively.

For the top SNPs reported for the combined analysis, there were no significant differences observed for effect size between men and women (Supplementary Material, Table S5). Any discrepancy between sexes appears to be due to difference in sample size rather than difference in effect.

DISCUSSION

We have performed the first comprehensive genome-wide association analysis of the human plasma glycome as represented by 46 glycosylation traits analysed in the plasma of 3533 individuals from four isolated European populations. The predominance of loci involved in fucosylation displaying the strongest association with glycan levels in plasma (13) was confirmed.

FUT8 is the fucosyltransferase responsible for the α1,6-fucosylation of the core *N*-acetyl-glucosamine (GlcNAc) structure of *N*-glycans (15). Glycan groups associated with the gene region containing *FUT8* contain fucose attached to their core, so the results are consistent with the known biological role of FUT8. In contrast, groups DG7, DG9 and DG12 include glycans containing antennary fucose, and FUC-A was derived as an overall measure of antennary fucosylation on biantennary glycans. *FUT6* encodes the enzyme fucosyltransferase VI, which was reported to be the key enzyme responsible for the α3-fucosylation of plasma proteins (14). The association of *FUT8*, *FUT6* and *FUT3* with *N*-glycan structures containing core and antennary fucosylation is supported by their known biological functions (16). A diagram illustrating where the

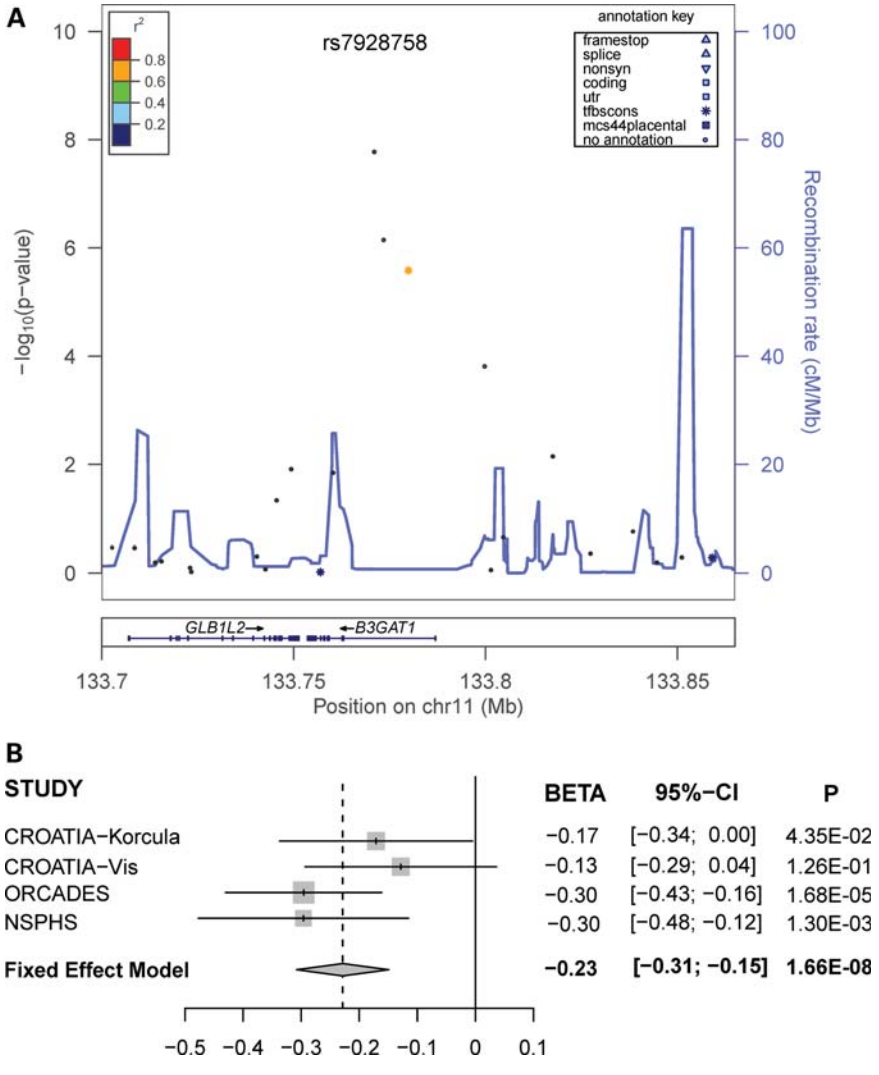

Figure 3. Significance plot (**A**) and forest plot (**B**) for the most significant SNP for regions of interest from the meta-analysis of DG13. (A) −Log 10 of the *P*-values are plotted against chromosome positions. The most significant SNP is labelled along with its *P*-value and information on SNP function (see annotation legend in the upper right corner). Gene information and regional LD (r^2) are also shown. (B) The estimate of the effect size of the best SNP for each individual population and the pool along with its standard error is shown, in which the size of the square for individual cohorts represents the proportion that it is contributing to the estimate of pooled effect size. The effect size presented is the β-coefficient, which represents a change in glycan levels measured in standard deviation units (adjusted for age, sex and PC) per copy of the allele modelled.

various fucosyltransferases act to add fucose in a generic *N*-glycan is presented in Figure 5.

In our previous study, *HNF1a* was associated with several glycan traits but the mechanism for this association was not obvious. Functional studies showed that HNF1a and HNF4a act to co-regulate the expression of most *fucosyltransferase* (*FUT3–11*) genes in HepG2 cells as well as gene expression levels of key enzymes needed for synthesis of GDP-fucose, the substrate for fucosyltransferases, thereby regulating both core and antennary fucosylation (13). Rare mutations within this gene cause Maturity Onset Diabetes of the Young 3 [MODY3 (MIM:600496)]. Common polymorphisms in *HNF1a* have been associated with variations of the plasma levels of C-reactive protein (17–22), low-density lipoprotein (23) and gamma-glutamyl transferase (24), and found to be susceptibility loci for type 2 diabetes (25,26), coronary heart disease (27) and pancreatic cancer (28). Although none of these associations reported the same top SNP as our study, none could be discounted from tagging the same region. There is strong support for the gamma-glutamyltransferase, coronary heart disease and C-reactive protein (rs1169310) loci to also influence *N*-glycan levels with the reported associated SNPs in complete linkage disequilibrium (LD) with our top SNP ($r^2 = 1$, $D' = 1$). The top SNP in the Voight *et al.* (25) type 2 diabetes study (rs7957197, $r^2 = 0.102$, $D' = 1$) is most likely not tagging the same pleiotropic signal but due to the very high D' value this cannot be completely ruled out without further investigation.

Three novel non-fucosylation-related associations were identified in this expanded analysis. The *MGAT5* gene codes for the enzyme mannosyl (α-1,6-)-glycoprotein β-1,6-*N*-acetyl-glucosaminyltransferase V (GnT-V), which adds GlcNAc residues to mannose in a β1,6 orientation on the antennary structure of *N*-glycans (Fig. 5). This is a decisive step

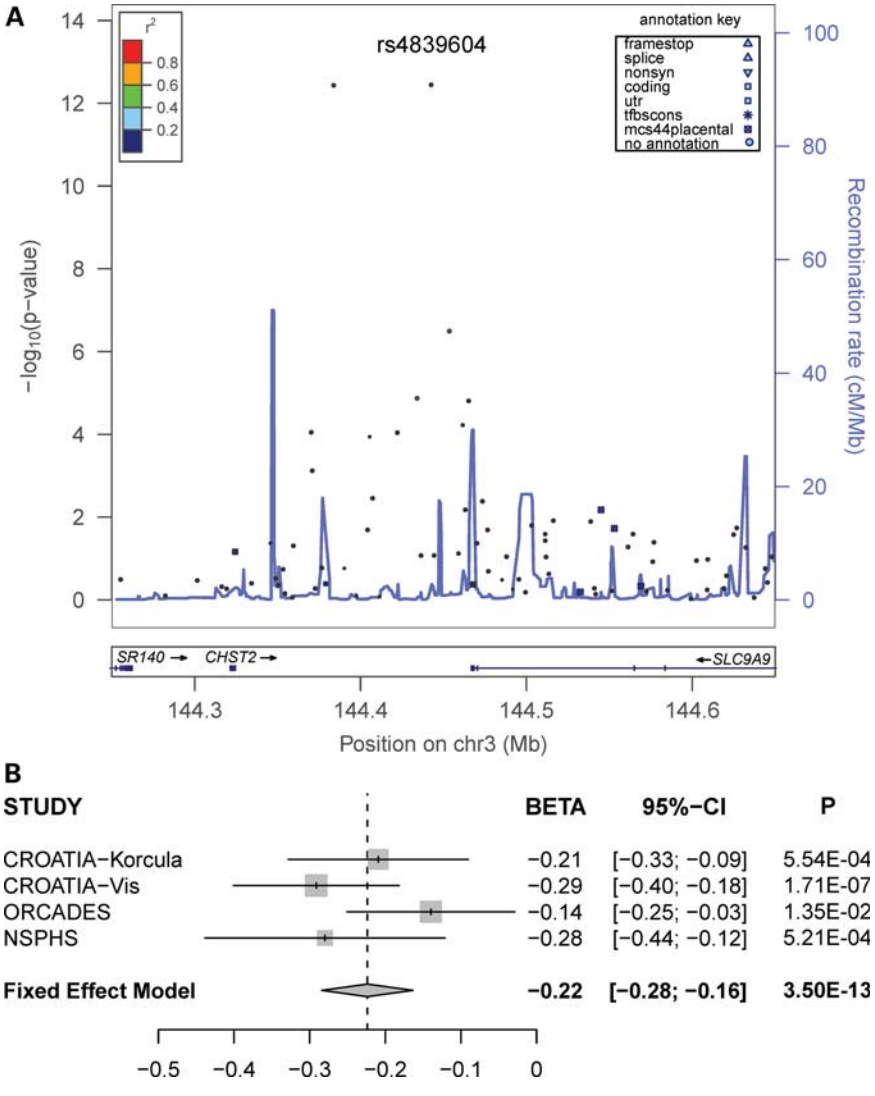

Figure 4. Significance plot (**A**) and forest plot (**B**) for the most significant SNP for regions of interest from the meta-analysis of tetrasialylated glycans. (A) −Log 10 of the *P*-values are plotted against chromosome positions. The most significant SNP is labelled along with its *P*-value and information on SNP function (see annotation legend in the upper right corner). Gene information and regional LD (r^2) are also shown. (B) The estimate of the effect size of the best SNP for each individual population and the pool along with its standard error is shown, in which the size of the square for individual cohorts represents the proportion that it is contributing to the estimate of pooled effect size. The effect size presented is the β-coefficient, which represents a change in glycan levels measured in standard deviation units (adjusted for age, sex and PC) per copy of the allele modelled.

in the generation of tetra-antennary glycans. The majority of structures contained within the DG11 peak are tetra-antennary glycans (29); therefore, the associations with DG11 and TA (a derived trait quantifying total tetra-antennary glycans) are biologically plausible. Tetra-antennary glycans are important regulators of membrane function (30) since they affect the half-life of numerous receptors on the cell membrane (31). This has important implications for many dynamic processes, from immunity to cancer progression and metastasis. MGAT5 synthesizes cell-surface ligands for galectins—proteins involved in the proliferation of T-cells and apoptosis. Loss of MGAT5 expression lowers the threshold needed for T-cell activation, and as Mgat5-deficient mice displayed several autoimmune phenotypes it was postulated that MGAT5 expression might be implicated in autoimmune disorders in humans (32). Recently, polymorphisms within MGAT5 have been tentatively associated with the severity of multiple sclerosis (33,34) in two small studies which may support this theory; however, the SNPs reported in these studies do not appear to be tagging the same LD blocks as our glycan-associated SNPs. Expression of MGAT5 is upregulated in oncogenic cells, and an increased production of galectin ligands on the cell surface allows the tumour cell to retain growth factors such as epidermal growth factor and transforming growth factor-β (35).

B3GAT1 is a member of the glucuronyltransferase gene family. This gene product functions as the key enzyme in a glucuronyl transfer reaction during the biosynthesis of the carbohydrate epitope HNK-1. It acts to add a glucuronic acid (GlcA) to the terminal *N*-acetyllactosamine (Lac) disaccharide to form the HNK-1 epitope precursor (36,37). The HNK-1 epitope is expressed on a subset of human

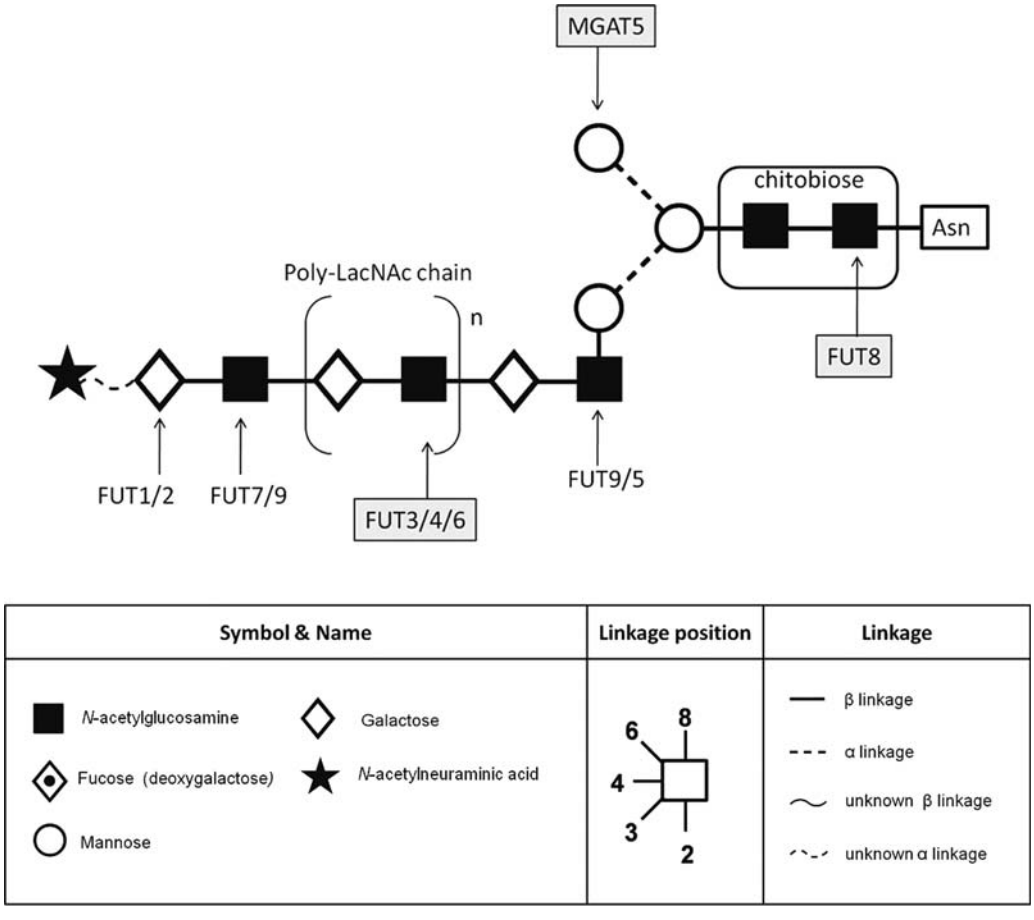

Figure 5. Pathways of core (tri-mannosyl-chitobiose) and antennary fucosylation. Genes highlighted indicate genome-wide significant genes identified from GWAS (*FUT6/FUT3*, *FUT8* and *MGAT5*). Structures are given using Oxford notation (62). Asn, asparagine residue on glycosylated protein.

lymphocytes, including natural killer cells, but it was not previously reported to exist on plasma proteins. We were able to show through mass spectrometry (MS) analysis that glucuronic acid is present on some glycans which make up the DG13 plasma glycan pool, explaining this association.

SLC9A9, solute carrier family 9 (sodium/hydrogen exchanger), is a proton pump which affects pH in the endosomal compartment (38). This gene was not previously linked to glycosylation, but it was recently reported that changes in Golgi pH can impair protein sialylation (39), thus the association between *SLC9A9* and tetrasialylated glycans makes biological sense. Sialic acids are found in cell secretions and are usually the terminal component of glycoproteins and glycolipids on the outer cell surface and therefore are involved in cell communication and defence. They act to shield recognition sites that may be antigenic in order to prevent autoimmunity but also function as ligands for many molecules such as hormones, inorganic cations and antibodies (40). Microorganisms are able to exploit the prominent role of sialic acids within the human body by either coating themselves with or binding to sialic acid in order to penetrate and infect the cell. A better understanding of processes and pathways underlying sialylation could lead to new avenues for the treatment of infection and disease (41). Polymorphisms in *SLC9A9* have recently been suggestively associated with attention-deficit hyperactive disorder [ADHD (MIM:143465)] (42,43). The first study reported only a gene-based *P*-value so it cannot be determined if the variant underlying the reported signal could be the same as for the glycosylation signal, whereas in the latter study the reported SNP tagging the disease signal is not in strong LD with our top SNP (rs1372288 and rs9810857, $r^2 = 0.063$, $D' = 0.575$; rs4839604 and rs9810857, $r^2 = 0.106$, $D' = 0.727$). Glycosylation analysis in ADHD patients also revealed difference in tetrasialylated glycans between children with ADHD and matching controls (44). Although disease causality cannot be ascertained by these data, the associations still provide a novel set of molecules which could act as clinical markers of disease.

Recent advances in high-throughput methods of analysing *N*-glycans have now made it possible to measure these traits in large cohorts. Analysing these traits with GWAS reveals several genome-wide significant associations illuminating the genetic control of distinct biological pathways, including fucosylation, sialylation and glucuronyl transfer, even in a modest sample size. Some of these biological processes are known to be altered in several disease states. For example, fucosylation of acute phase proteins is modified in many diseases, such as acute inflammation (45,46), rheumatoid arthritis (47) and diabetes (48), and changes in the levels of fucosylated glycans have been shown to be associated with several important pathological processes, including cancer and inflammation

(49). Additionally, the finding that loci associated with disease (e.g. *SLC9A9* in ADHD or *HNF1a* in MODY) are modulating various glycan species offers novel insight into disease mechanisms and pathways. Variation in the glycosylation of plasma proteins caused by the polymorphisms identified here could be a predisposing or prognostic factor in numerous diseases and warrants further examination of these effects in plasma samples from specific disease cohorts.

MATERIALS AND METHODS

Study participants

All population studies recruited adult individuals within a community irrespective of any specific phenotype. Fasting blood samples were collected, biochemical and physiological measurements taken and questionnaire data for medical history as well as lifestyle and environmental exposures collected following similar protocols. All studies conformed to the ethical guidelines of the 1975 Declaration of Helsinki and were approved by appropriate ethics boards, with all participants signing informed consent prior to participation.

The CROATIA-Vis study includes 1008 Croatians, aged 18–93 years, who were recruited from the villages of Vis and Komiza on the Dalmatian Island of Vis during 2003 and 2004 within a larger genetic epidemiology program (50,51).

The CROATIA-Korcula study includes 969 Croatians between the ages of 18 and 98 (52). The field work was performed in 2007 and 2008 in the eastern part of the island, targeting healthy volunteers from the town of Korčula and the villages of Lumbarda, Žrnovo and Račišće.

ORCADES was performed in the Scottish archipelago of Orkney, which collected data between 2005 and 2011 (53). Data for 889 participants aged 18–100 years from a subgroup of 10 islands were used for this analysis.

NSPHS is a family-based population study including a comprehensive health investigation and collection of data on family structure, lifestyle, diet, medical history and samples for laboratory analyses from peoples living in the north of Sweden (54). Data were available from 700 participants aged 14–91 years.

DNA samples were genotyped according to the manufacturer's instructions on Illumina Infinium SNP bead microarrays (HumanHap300v1 for CROATIA-Vis, HumanHap300v2 for ORCADES and NSPHS and HumanCNV370v1 for CROATIA-Korcula). Genotypes were determined using the Illumina BeadStudio software. Genotyping was successfully completed on 991 individuals from CROATIA-Vis, 953 from CROATIA-Korcula, 889 from ORCADES and 700 from NSPHS.

N-glycan measurement and trait description

Plasma *N*-glycans were measured using hydrophilic interaction HPLC (HILIC) and weak anion exchange (WAX) HPLC as previously described (10,50,55). Glycans were identified on the basis of their elution positions, measured in glucose units and then compared with reference values in NIBRT's 'GlycoBase v3.0' database for structure assignment (50). Structures were confirmed by exoglycosidase digestion as previously reported (29). These methods have shown to be robust and reproducible across measurement centres (13,29). Chromatography yielded a total of 33 measured traits. From the HPLC analysis, undigested *N*-glycans were separated into 16 structurally related groups of glycans, referred to as glycan peak (GP)1–16. Desialylated 2AB-labelled *N*-glycans were also separated into 13 structurally related groups of glycans, referred to as DG1–DG13. WAX chromatography resulted in four peaks: monosialylated, disialylated, trisialylated and tetrasialylated structures. The amount of *N*-glycans measured in each of these groups was quantified relatively, as a proportion of the total plasma *N*-glycome. An additional 13 parameters were gained by calculating some structural determinants from measured traits. A description of each glycan trait is presented in Table 2 along with equations for the additional 13 derived traits.

Genotype and phenotype quality control

Genotyping quality control was performed using the same procedures for all cohorts. Individuals with a call rate $<97\%$ were removed as well as SNPs with a call rate $<98\%$ (95% for CROATIA-Vis), minor allele frequency <0.02 or Hardy–Weinberg equilibrium *P*-value less than 1×10^{-10}. In total, 924 individuals passed all quality control thresholds from CROATIA-Vis, 898 from CROATIA-Korcula, 889 from ORCADES and 656 from NSPHS. The number of participants available for meta-analysis prior to phenotype quality control was 3367 (1941 women, 1426 men).

Extreme outliers (those with values more than three times the interquartile distances away from either the 75th or the 25th percentile values) were removed for each glycan measure to eliminate individuals not representative of normal variation within the population.

Genome-wide association analysis

Genome-wide association analysis was performed for each population and then combined using an inverse-variance weighted meta-analysis for all traits. Each trait was adjusted for sex, age and the first three PC obtained from the population-specific identity-by-state (IBS)-derived distance matrix (described below) and the residuals transformed to ensure their normal distribution using quantile normalization. Sex-specific analyses were adjusted for age and PC only. The residuals expressed as *z*-scores were used for association analysis. The 'mmscore' function of the GenABEL package (56) for the R statistical software was used for the association test under an additive model. This score test for family-based association takes into account pedigree structure and allows unbiased estimations of SNP allelic effect when relatedness is present between examinees (57). The relationship matrix used in this analysis was generated by the 'ibs()' function of GenABEL (using weight = 'freq' option), which utilizes IBS genotype sharing to estimate the realized pairwise kinship coefficient. All lambda values for the population-specific analyses were <1.05, showing that this method efficiently accounts for family structure. Inverse-variance weighted meta-analysis was performed using the MetABEL package (56) for R. An association was considered statistically significant at the genome-wide level if the *P*-value for an individual SNP was

Table 2. Description of plasma *N*-glycan traits analysed

Trait code	Trait	Trait description
GP1	Glycan Peak 1	Relative percentage of specific peak area/all peaks area from the HILIC profile
GP2	Glycan Peak 2	
GP3	Glycan Peak 3	
GP4	Glycan Peak 4	
GP5	Glycan Peak 5	
GP6	Glycan Peak 6	
GP7	Glycan Peak 7	
GP8	Glycan Peak 8	
GP9	Glycan Peak 9	
GP10	Glycan Peak 10	
GP11	Glycan Peak 11	
GP12	Glycan Peak 12	
GP13	Glycan Peak 13	
GP14	Glycan Peak 14	
GP15	Glycan Peak 15	
GP16	Glycan Peak 16	
DG1	Desialylated Peak 1	Relative percentage of specific peak area/all peaks area from the HILIC profile after sialidase treatment
DG2	Desialylated Peak 2	
DG3	Desialylated Peak 3	
DG4	Desialylated Peak 4	
DG5	Desialylated Peak 5	
DG6	Desialylated Peak 6	
DG7	Desialylated Peak 7	
DG8	Desialylated Peak 8	
DG9	Desialylated Peak 9	
DG10	Desialylated Peak 10	
DG11	Desialylated Peak 11	
DG12	Desialylated Peak 12	
DG13	Desialylated Peak 13	
Monosialylated	Percent monosialylated	Obtained from WAX analysis
Disialylated	Percent disialylated	
Trisialylated	Percent trisialylated	
Tetrasialylated	Percent tetrasialylated	
C-FUC	Core fucosylated glycans	DG6/(DG5 + DG6)
A-FUC	Antennary fucosylated glycans	DG7/(DG5 + DG7)
A2	Biantennary nongalactosylated glycans	(GP1 + DG1)/2
BA	Biantennary glycans	DG1 + DG2 + DG3 + DG4 + DG5 + DG6 + DG7
BAMS	Monosialylated biantennary glycans	(GP7 + GP8)/(DG5 + DG6 + DG7)
BADS	Disialylated biantennary glycans	(GP9 + GP10 + GP11)/(DG5 + DG6 + DG7)
TRIA	Triantennary glycans	DG8 + DG9 + DG10
TA	Tetra-antennary glycans	DG11 + DG12 + DG13
G0	Nongalactosylated glycans	DG1 + DG2
G1	Monogalactosylated glycans	DG3 + DG4
G2	Digalactosylated glycans	DG5 + DG6 + DG7
G3	Trigalactosylated glycans	GP12 + GP13 + GP14
G4	Tetragalactosylated glycans	GP15 + GP16

less than 5×10^{-8}. The effect of the most significant SNP in each gene region expressed as a percentage of the variance explained was calculated for each glycan trait adjusted for sex, age and first three PC in each cohort individually using the 'polygenic' function of the GenABEL package for R.

Result interpretation

Regions of association were visualized using the web-based software LocusZoom (58) to display the LD of the region. A preliminary literature search for each associated gene was performed using GWAS Catalog (59) for associations reaching genome-wide significance within the study or reasonably suggestive if sample size was large ($\geq$10 000). Additional associations were reported if they were replicated by another population either within the original publication or in another study. SNAP (60) was used to look at LD between SNP associations from the literature and this study. 1000 Genomes Pilot 1 data from the CEU population was used to calculate LD, and SNPs were accepted to be tagging the same region if $D' = 1$ and $r^2 > 0.8$.

Mass spectrometry

Nano-LC-ESI-MS/MS. MS analysis of the collected glycan fractions was performed using an Ultimate 3000 nano-LC system (Dionex/LC Packings, Amsterdam, The Netherlands) equipped with a reverse-phase trap column (C_{18} PepMap 100 Å, 5 μm, 300 μm × 5 mm; Dionex/LC Packings) and a nano column (C_{18} PepMap 100 Å, 3 μm, 75 μm × 150 mm; Dionex/LC Packings). The column was equilibrated at room temperature with eluent A (0.1% formic acid in water) at a

flow rate of 300 nl/min. After injection of the samples, a gradient was applied to 25% eluent B (95% acetonitrile) in 15 min and to 70% eluent B at 25 min, followed by an isocratic elution with 70% eluent B for 5 min. The eluate was monitored by UV absorption at 214 nm. The LC system was coupled via an online nanospray source to an Esquire HCTultra ESI-IT-MS (Bruker Daltonics, Bremen, Germany) operated in the positive ion mode. For electrospray (1100–1250 V), stainless steel capillaries with an inner diameter of 30 μm (Proxeon, Odense, Denmark) were used. The solvent was evaporated at 170°C, employing a nitrogen stream of 7 l/min. Ions from *m*/*z* 500 to 1800 were registered. Automatic fragment ion analysis was enabled, resulting in MS/MS spectra of the most abundant ions in the MS spectra. Glycan structures were assigned using GlycoWorkbench (61).

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

ACKNOWLEDGEMENTS

The CROATIA-Vis and CROATIA-Korcula studies would like to acknowledge the invaluable contributions of the recruitment team (including those from the Institute of Anthropological Research in Zagreb) in Vis and Korcula, the administrative teams in Croatia and Edinburgh and the people of Vis and Korcula. ORCADES would like to acknowledge the invaluable contributions of Lorraine Anderson, the research nurses in Orkney and the administrative team in Edinburgh. SNP genotyping of the CROATIA-Vis samples and DNA extraction for ORCADES were carried out by the Genetics Core Laboratory at the Wellcome Trust Clinical Research Facility, WGH, Edinburgh, UK, SNP genotyping for CROATIA-Korcula and ORCADES was performed by Helmholtz Zentrum München, GmbH, Neuherberg, Germany.

Conflict of Interest statement. None declared.

FUNDING

The CROATIA-Vis and CROATIA-Korcula studies in the Croatian islands of Vis and Korcula were supported by grants from the Medical Research Council (UK), the Ministry of Science, Education and Sport of the Republic of Croatia (grant number 108–1080315–0302) and the European Union framework program 6 European Special Populations Research Network project (contract LSHG-CT-2006–018947). ORCADES was supported by the Chief Scientist Office of the Scottish Government, the Royal Society and the European Union framework program 6 European Special Populations Research Network project (contract LSHG-CT-2006–018947). Glycome analysis was supported by the Croatian Ministry of Science, Education and Sport (grant number 309–0061194–2023), the Croatian Science Foundation (grant number 04–47) and the European Commission (grants EuroGlycoArrays and GlycoBioM). B.A. and I.R. acknowledge the European Union framework program 7 EuroGlycoArrays ITN (grant number 215536) for funding.

REFERENCES

1. Marek, K.W., Vijay, I.K. and Marth, J.D. (1999) A recessive deletion in the GlcNAc-1-phosphotransferase gene results in periimplantation embryonic lethality. *Glycobiology*, **9**, 1263–1271.
2. Freeze, H.H. (2006) Genetic defects in the human glycome. *Nat. Rev. Genet.*, **7**, 537–551.
3. Jaeken, J. (2003) Congenital disorders of glycosylation (CDG): It's all in it! *J. Inherit. Metab. Dis.*, **26**, 99–118.
4. Ohtsubo, K. and Marth, J.D. (2006) Glycosylation in cellular mechanisms of health and disease. *Cell*, **126**, 855–867.
5. Hart, G.W., Housley, M.P. and Slawson, C. (2007) Cycling of *O*-linked beta-*N*-acetylglucosamine on nucleocytoplasmic proteins. *Nature*, **446**, 1017–1022.
6. Crocker, P.R., Paulson, J.C. and Varki, A. (2007) Siglecs and their roles in the immune system. *Nat. Rev. Immunol.*, **7**, 255–266.
7. Marth, J.D. and Grewal, P.K. (2008) Mammalian glycosylation in immunity. *Nat. Rev. Immunol.*, **8**, 874–887.
8. Varki, A. (1993) Biological roles of oligosaccharides – all of the theories are correct. *Glycobiology*, **3**, 97–130.
9. Cummings, R.D. (2009) The repertoire of glycan determinants in the human glycome. *Mol. Biosyst.*, **5**, 1087–1104.
10. Royle, L., Campbell, M.P., Radcliffe, C.M., White, D.M., Harvey, D.J., Abrahams, J.L., Kim, Y.G., Henry, G.W., Shadick, N.A., Weinblatt, M.E. *et al.* (2008) HPLC-based analysis of serum *N*-glycans on a 96-well plate platform with dedicated database software. *Anal. Biochem.*, **376**, 1–12.
11. Rudd, P.M., Rudan, I. and Wright, A.F. (2009) High-throughput glycome analysis is set to join high-throughput genomics. *J. Proteome Res.*, **8**, 1105.
12. Knezevic, A., Gornik, O., Polasek, O., Pucic, M., Redzic, I., Novokmet, M., Rudd, P.M., Wright, A.F., Campbell, H., Rudan, I. and Lauc, G. (2010) Effects of aging, body mass index, plasma lipid profiles, and smoking on human plasma *N*-glycans. *Glycobiology*, **20**, 959–969.
13. Lauc, G., Essafi, A., Huffman, J.E., Hayward, C., Knezevic, A., Kattla, J.J., Polasek, O., Gornik, O., Vitart, V., Abrahams, J.L. *et al.* (2010) Genomics meets glycomics - the first GWAS study of human *N*-glycome identifies HNF1alpha as a master regulator of plasma protein fucosylation. *PLoS Genet.*, **6**, e1001256.
14. Brinkman-Van der Linden, E.C.M., Mollicone, R., Oriol, R., Larson, G., Van den Eijnden, D.H. and Van Dijk, W. (1996) A missense mutation in the FUT6 gene results in total absence of alpha 3-fucosylation of human alpha(1)-acid glycoprotein. *J. Biol. Chem.*, **271**, 14492–14495.
15. Taniguchi, N., Honke, K. and Fukuda, M. (2002) *Handbook of Glycosyltransferases and Related Genes*. Springer-Verlag, Tokyo.
16. Ma, B., Simala-Grant, J.L. and Taylor, D.E. (2006) Fucosylation in prokaryotes and eukaryotes. *Glycobiology*, **16**, 158R–184R.
17. Dehghan, A., Dupuis, J., Barbalic, M., Bis, J.C., Eiriksdottir, G., Lu, C., Pellikka, N., Wallaschofski, H., Kettunen, J., Henneman, P. *et al.* (2011) Meta-analysis of genome-wide association studies in >80 000 subjects identifies multiple loci for C-reactive protein levels. *Circulation*, **123**, 731–738.
18. Elliott, P., Chambers, J.C., Zhang, W.H., Clarke, R., Hopewell, J.C., Peden, J.F., Erdmann, J., Braund, P., Engert, J.C., Bennett, D. *et al.* (2009) Genetic loci associated with C-reactive protein levels and risk of coronary heart disease. *J. Am. Med. Assoc.*, **302**, 37–48.
19. Ridker, P.M., Pare, G., Parker, A., Zee, R.Y.L., Danik, J.S., Buring, J.E., Kwiatkowski, D., Cook, N.R., Miletich, J.P. and Chasman, D.I. (2008) Loci related to metabolic-syndrome pathways including LEPR, HNF1A, IL6R, and GCKR associate with plasma C-reactive protein: the women's genome health study. *Am. J. Hum. Genet.*, **82**, 1185–1192.
20. Reiner, A.P., Barber, M.J., Guan, Y., Ridker, P.M., Lange, L.A., Chasman, D.I., Walston, J.D., Cooper, G.M., Jenny, N.S., Rieder, M.J. *et al.* (2008) Polymorphisms of the HNF1A gene encoding hepatocyte nuclear factor-1 alpha are associated with C-reactive protein. *Am. J. Hum. Genet.*, **82**, 1193–1201.
21. Okada, Y., Takahashi, A., Ohmiya, H., Kumasaka, N., Kamatani, Y., Hosono, N., Tsunoda, T., Matsuda, K., Tanaka, T., Kubo, M. *et al.* (2011) Genome-wide association study for C-reactive protein levels identified pleiotropic associations in the IL6 locus. *Hum. Mol. Genet.*, **20**, 1224–1231.
22. Wu, Y., McDade, T.W., Kuzawa, C.W., Borja, J., Li, Y., Adair, L.S., Mohlke, K.L. and Lange, L.A. (2011) Genome-wide association with C-reactive protein levels in CLHNS: evidence for the CRP and HNF1A loci and their interaction with exposure to a pathogenic environment. *Inflammation*, doi: 10.1007/s10753-011-9348-y.

23. Kathiresan, S., Willer, C.J., Peloso, G.M., Demissie, S., Musunuru, K., Schadt, E.E., Kaplan, L., Bennett, D., Li, Y., Tanaka, T. *et al.* (2009) Common variants at 30 loci contribute to polygenic dyslipidemia. *Nat. Genet.*, **41**, 56–65.
24. Yuan, X., Waterworth, D., Perry, J.R.B., Lim, N., Song, K., Chambers, J.C., Zhang, W., Vollenweider, P., Stirnadel, H., Johnson, T. *et al.* (2008) Population-based genome-wide association studies reveal six loci influencing plasma levels of liver enzymes. *Am. J. Hum. Genet.*, **83**, 520–528.
25. Voight, B.F., Scott, L.J., Steinthorsdottir, V., Morris, A.P., Dina, C., Welch, R.P., Zeggini, E., Huth, C., Aulchenko, Y.S., Thorleifsson, G. *et al.* (2010) Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat. Genet.*, **42**, 579–589.
26. Parra, E.J., Below, J.E., Krithika, S., Valladares, A., Barta, J.L., Cox, N.J., Hanis, C.L., Wacher, N., Garcia-Mena, J., Hu, P. *et al.* (2011) Genome-wide association study of type 2 diabetes in a sample from Mexico City and a meta-analysis of a Mexican-American sample from Starr County, Texas. *Diabetologia*, **54**, 2038–2046.
27. Erdmann, J., Grosshennig, A., Braund, P.S., Konig, I.R., Hengstenberg, C., Hall, A.S., Linsel-Nitschke, P., Kathiresan, S., Wright, B., Tregouet, D.A. *et al.* (2009) New susceptibility locus for coronary artery disease on chromosome 3q22.3. *Nat. Genet.*, **41**, 280–282.
28. Pierce, B.L. and Ahsan, H. (2011) Genome-wide 'pleiotropy scan' identifies HNF1A region as a novel pancreatic cancer susceptibility locus. *Cancer Res.*, **71**, 4352–4358.
29. Knezevic, A., Polasek, O., Gornik, O., Rudan, I., Campbell, H., Hayward, C., Wright, A., Kolcic, I., O'Donoghue, N., Bones, J. *et al.* (2009) Variability, heritability and environmental determinants of human plasma *N*-glycome. *J. Proteome Res.*, **8**, 694–701.
30. Dennis, J.W., Lau, K.S., Demetriou, M. and Nabi, I.R. (2009) Adaptive regulation at the cell surface by N-glycosylation. *Traffic*, **10**, 1569–1578.
31. Partridge, E.A., Le, R.C., Di Guglielmo, G.M., Pawling, J., Cheung, P., Granovsky, M., Nabi, I.R., Wrana, J.L. and Dennis, J.W. (2004) Regulation of cytokine receptors by Golgi *N*-glycan processing and endocytosis. *Science*, **306**, 120–124.
32. Demetriou, M., Granovsky, M., Quaggin, S. and Dennis, J.W. (2001) Negative regulation of T-cell activation and autoimmunity by Mgat5 N-glycosylation. *Nature*, **409**, 733–739.
33. Brynedal, B., Wojcik, J., Esposito, F., Debailleul, V., Yaouanq, J., Martinelli-Boneschi, F., Edan, G., Comi, G., Hillert, J. and Abderrahim, H. (2010) MGAT5 alters the severity of multiple sclerosis. *J. Neuroimmunol.*, **220**, 120–124.
34. Esposito, F., Wojcik, J., Rodegher, M., Radaelli, M., Moiola, L., Ghezzi, A., Capra, R., Brambilla, P., Sorosina, M., Giacalone, G. *et al.* (2011) MGAT5 and disease severity in progressive multiple sclerosis. *J. Neuroimmunol.*, **230**, 143–147.
35. Lau, K.S. and Dennis, J.W. (2008) *N*-Glycans in cancer progression. *Glycobiology*, **18**, 750–760.
36. Oka, S., Terayama, K., Kawashima, C. and Kawasaki, T. (1992) A novel glucuronyltransferase in nervous system presumably associated with the biosynthesis of HNK-1 carbohydrate epitope on glycoproteins. *J. Biol. Chem.*, **267**, 22711–22714.
37. Mitsumoto, Y., Oka, S., Sakuma, H., Inazawa, J. and Kawasaki, T. (2000) Cloning and chromosomal mapping of human glucuronyltransferase involved in biosynthesis of the HNK-1 carbohydrate epitope. *Genomics*, **65**, 166–173.
38. Roxrud, I., Raiborg, C., Gilfillan, G.D., Stromme, P. and Stenmark, H. (2009) Dual degradation mechanisms ensure disposal of NHE6 mutant protein associated with neurological disease. *Exp. Cell Res.*, **315**, 3014–3027.
39. Rivinoja, A., Hassinen, A., Kokkonen, N., Kauppila, A. and Kellokumpu, S. (2009) Elevated Golgi pH impairs terminal N-glycosylation by inducing mislocalization of Golgi glycosyltransferases. *J. Cell. Physiol.*, **220**, 144–154.
40. Schauer, R. (2009) Sialic acids as regulators of molecular and cellular interactions. *Curr. Opin. Struct. Biol.*, **19**, 507–514.
41. Varki, A. (2008) Sialic acids in human health and disease. *Trends Mol. Med.*, **14**, 351–360.
42. Brookes, K., Xu, X., Chen, W., Zhou, K., Neale, B., Lowe, N., Anney, R., Franke, B., Gill, M., Ebstein, R. *et al.* (2006) The analysis of 51 genes in DSM-IV combined type attention deficit hyperactivity disorder: association signals in DRD4, DAT1 and 16 other genes. *Mol. Psychiatry*, **11**, 934–953.
43. Mick, E., Todorov, A., Smalley, S., Hu, X., Loo, S., Todd, R.D., Biederman, J., Byrne, D., Dechairo, B., Guiney, A. *et al.* (2010) Family-based genome-wide association scan of attention-deficit/hyperactivity disorder. *J. Am. Acad. Child Adolesc. Psychiatry*, **49**, 898–905.
44. Pivac, N., Knezevic, A., Gornik, O., Pucic, M., Igl, W., Peeters, H., Crepel, A., Steyaert, J., Novokmet, M., Redzic, I. *et al.* (2011) Human plasma glycome in attention-deficit hyperactivity disorder and autism spectrum disorders. *Mol. Cell. Proteomics*, **10**, M110.
45. Brinkman-van der Linden, E.C., de Haan, P.F., Havenaar, E.C. and van Dijk, W. (1998) Inflammation-induced expression of sialyl LewisX is not restricted to alpha1-acid glycoprotein but also occurs to a lesser extent on alpha1-antichymotrypsin and haptoglobin. *Glycoconj. J.*, **15**, 177–182.
46. Higai, K., Aoki, Y., Azuma, Y. and Matsumoto, K. (2005) Glycosylation of site-specific glycans of alpha1-acid glycoprotein and alterations in acute and chronic inflammation. *Biochim. Biophys. Acta*, **1725**, 128–135.
47. Higai, K., Azuma, Y., Aoki, Y. and Matsumoto, K. (2003) Altered glycosylation of alpha1-acid glycoprotein in patients with inflammation and diabetes mellitus. *Clin. Chim. Acta*, **329**, 117–125.
48. Poland, D.C., Schalkwijk, C.G., Stehouwer, C.D., Koeleman, C.A., van het, H.B. and van Dijk, W. (2001) Increased alpha3-fucosylation of alpha1-acid glycoprotein in Type I diabetic patients is related to vascular function. *Glycoconj. J.*, **18**, 261–268.
49. Peracaula, R., Sarrats, A. and Rudd, P.M. (2010) Liver proteins as sensor of human malignancies and inflammation. *Proteomics Clin. Appl.*, **4**, 426–431.
50. Campbell, M.P., Royle, L., Radcliffe, C.M., Dwek, R.A. and Rudd, P.M. (2008) GlycoBase and autoGU: tools for HPLC-based glycan analysis. *Bioinformatics*, **24**, 1214–1216.
51. Vitart, V., Rudan, I., Hayward, C., Gray, N.K., Floyd, J., Palmer, C.N., Knott, S.A., Kolcic, I., Polasek, O., Graessler, J. *et al.* (2008) SLC2A9 is a newly identified urate transporter influencing serum urate concentration, urate excretion and gout. *Nat. Genet.*, **40**, 437–442.
52. Zemunik, T., Boban, M., Lauc, G., Jankovic, S., Rotim, K., Vatavuk, Z., Bencic, G., Dogas, Z., Boraska, V., Torlak, V. *et al.* (2009) Genome-wide association study of biochemical traits in Korcula Island, Croatia. *Croat. Med. J.*, **50**, 23–33.
53. McQuillan, R., Leutenegger, A.L., Abdel-Rahman, R., Franklin, C.S., Pericic, M., Barac-Lauc, L., Smolej-Narancic, N., Janicijevic, B., Polasek, O., Tenesa, A. *et al.* (2008) Runs of homozygosity in European populations. *Am. J. Hum. Genet.*, **83**, 359–372.
54. Kolz, M., Johnson, T., Sanna, S., Teumer, A., Vitart, V., Perola, M., Mangino, M., Albrecht, E., Wallace, C., Farrall, M. *et al.* (2009) Meta-analysis of 28,141 individuals identifies common variants within five new loci that influence uric acid concentrations. *PLoS Genet.*, **5**, e1000504.
55. Royle, L., Radcliffe, C.M., Dwek, R.A. and Rudd, P.M. (2006) Detailed structural analysis of *N*-glycans released from glycoproteins in SDS-PAGE gel bands using HPLC combined with exoglycosidase array digestions. *Methods Mol. Biol.*, **347**, 125–143.
56. Aulchenko, Y.S., Ripke, S., Isaacs, A. and Van Duijn, C.M. (2007) GenABEL: an R library for genorne-wide association analysis. *Bioinformatics*, **23**, 1294–1296.
57. Chen, W.M. and Abecasis, G.R. (2007) Family-based association tests for genomewide association scans. *Am. J. Hum. Genet.*, **81**, 913–926.
58. Pruim, R.J., Welch, R.P., Sanna, S., Teslovich, T.M., Chines, P.S., Gliedt, T.P., Boehnke, M., Abecasis, G.R. and Willer, C.J. (2010) LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics*, **26**, 2336–2337.
59. Hindorff, L.A., Sethupathy, P., Junkins, H.A., Ramos, E.M., Mehta, J.P., Collins, F.S. and Manolio, T.A. (2009) Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl Acad. Sci. USA*, **106**, 9362–9367.
60. Johnson, A.D., Handsaker, R.E., Pulit, S.L., Nizzari, M.M., O'Donnell, C.J. and de Bakker, P.I.W. (2008) SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics*, **24**, 2938–2939.
61. Ceroni, A., Maass, K., Geyer, H., Geyer, R., Dell, A. and Haslam, S.M. (2008) GlycoWorkbench: a tool for the computer-assisted annotation of mass spectra of glycans. *J. Proteome Res.*, **7**, 1650–1659.
62. Harvey, D.J., Merry, A.H., Royle, L., Campbell, M.P., Dwek, R.A. and Rudd, P.M. (2009) Proposal for a standard system for drawing structural diagrams of N- and O-linked carbohydrates and related compounds. *Proteomics*, **9**, 3796–3801.