BRIEF COMMUNICATION

Genome-wide association study identifies *FUT8* and *ESR2* as co-regulators of a bi-antennary N-linked glycan A2 ($GlcNAc_2Man_3GlcNAc_2$) in human plasma proteins

Gordan Lauc[1,2], Jennifer Huffman[3], Caroline Hayward[3], Ana Knezevic[1], Ozren Polasek[4], Olga Gornik[1], Veronique Vitart[3], Ivana Kolcic[4], Zrinka Biloglav[4], Lina Zgaga[4], Nicholas D. Hastie[3], Alan F. Wright[3], Harry Campbell[5], Pauline M. Rudd[6], Igor Rudan[5,7,8]

[1] *Department of Biochemistry and Molecular Biology, University of Zagreb Faculty of Pharmacy and Biochemistry, Zagreb, Croatia.*
[2] *Glycobiology Laboratory, Genos Ltd, Planinska 1, 10000 Zagreb, Croatia*
[3] *MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, Western General Hospital, Crewe Road, Edinburgh, UK*
[4] *Andrija Štampar School of Public Health, University of Zagreb Medical School, Zagreb, Croatia*
[5] *Department of Public Health Sciences, The University of Edinburgh Medical School, Teviot Place, Edinburgh, UK*
[6] *NIBRT, Dublin-Oxford Glycobiology Lab., Conway Institute, University College Dublin, Belfield, Dublin 4, Ireland*
[7] *Croatian Centre for Global Health, University of Split Medical School, Split, Croatia*
[8] *Institute for Clinical Medical Research, University Hospital "Sestre milosrdnice", Zagreb, Croatia*

To whom correspondence should be addressed:

Igor Rudan, MD, DSc, PhD, MPH
Department of Public Health Sciences,
The University of Edinburgh Medical School,
Teviot Place, Edinburgh EH8 9AG, Scotland, UK
E-mail Igor.Rudan@ed.ac.uk

Abstract

HPLC analysis of N-glycans quantified levels of the biantennary glycan (A2) in plasma proteins of 924 individuals. Subsequent genome-wide association study (GWAS) using 317,503 single nucleotide polymorphysms (SNP) identified two genetic loci influencing variation in A2: FUT 8 and ESR2. We demonstrate that human glycans are amenable to GWAS and their genetic regulation shows sex-specific effects with *FUT 8* variants explaining 17.3% of the variance in pre-menopausal women, while *ESR2* variants explained 6.0% of the variance in post-menopausal women.

Main text

Glycosylation, which involves the addition of sugar chains to proteins and lipids, is the most complex and abundant post-translational modification (**1**). This process is not directly template driven and gives rise to the extensive complexity of the glycoproteome, estimated to be several orders of magnitude more complex than the proteome itself (**2**). In eukaryotes, glycosylation occurs through 11 biosynthetic pathways (**3**). One of these pathways is N-glycosylation, which occurs when a block of 14 sugars (the dolichol phosphate precursor) is transferred co-translationally to specific Asn residues in newly synthesized polypeptides in the endoplasmic reticulum. The resulting N-glycans are subjected to extensive modification as the glycoproteins mature and move via the Golgi complex to their intra- and extra-cellular destinations (**4**).

N-glycosylation is essential for multicellular life and its complete absence is embryonically lethal (**5**). Genetic defects that affect protein glycosylation through biosynthetic or degradation pathways are relevant to human health and are at the root of at least 30 human diseases (**2,6**). A surprisingly high proportion of disease-causing mis-sense mutations result in gains of glycosylation sites (**7,8**). Population variability in terminal glycans is also common (e.g., ABO blood groups) and contributes to the protein heterogeneity that can be advantageous for evading pathogens and adapting to a changing environments (**9**). However, due to experimental limitations in quantifying glycans in complex biological samples, understanding of the genetic regulation of glycosylation is still very limited (**2**).

Until recently, detailed quantitative analysis of total plasma glycans was generally performed on a limited number of samples due to the technological challenges (**10**). However, using a newly developed 96 well plate platform we have demonstrated that human plasma N-glycans can be separated into chromatographic peaks and reliably quantified using high-performance liquid chromatography (HPLC) (**11,12**). This represents a major technological advance, which was required to investigate the genetic regulation and biological role of glycan structures and for bringing glycomics into line with genomics, proteomics and metabolomics. We analysed the distribution of plasma glycans in a sample of 924 individuals recruited as part of a larger study of several hundred disease-related quantitative traits in a Croatian isolated population (**12,13**). Firstly, we found unusual biological variability at the population level with the median ratio of minimal to maximal values of 6.2 (from data on 33 chromatographic classes of N glycans (**12**)) and significant age- and gender-specific effects were also present. Heritability estimates for individual glycans varied widely, ranging from very low to very high. Environmental determinants were also detected, including diet and smoking (**12**).

In parallel to these developments in high-throughput glycome analysis, high-throughput genome analysis has brought unprecedented successes in gene identification in recent years (**14**). Genome-wide association studies are now established as a reliable and powerful tool to identify genetic variants underlying biological variation in quantitative traits and complex human diseases (**15**). This paper represents the first attempt to investigate whether a GWAS approach can be applied to the study of genetic regulation of individual glycan in human plasma quantified by HPLC.

In the HPLC method, only one glycan structure (A2, GlcNAc2Man3GlcNAc2) was individually quantified, while all others were clustered into groups that share similar structures and require further enzymatic treatment to separate them (**12**). Glycan A2 mostly

originates from immunoglobulin G (IgG) and differences in glycans attached to IgG glycoforms potentially affect disease susceptibility through effects on immune function (**16,17**). Glycan A2 showed substantial heritability in our study sample of 924 individuals residing in Croatian island isolates (h2=0.48; SE(h2) = 0.11). After initial HPLC quantification of glycan A2 we performed sialidase digestion to improve measurement precision. Trait concentrations were then $\log_{10}$-transformed, standardized by sex and age and z-transformed prior to genetic analysis. We successfully performed a genome-wide association study using 317,503 single nucleotide polymorphisms (SNPs) in 924 individuals from the Croatian isolate (**12**). An association was considered statistically significant at the genome-wide level if the p-value for an individual SNP was less that 1.7×10^{-7} (based on Bonferroni correction to account for multiple testing). For details of the methods used see **Supplementary Methods**.

The genome-wide association study identified a "peak" on chromosome 14 (**Figure 1**), which included 6 single nucleotide polymorphisms (SNPs) showing association with glycan A2 and genome-wide significance. The most significant association was found with single nucleotide polymorphism rs7161123 located in intron 2 of the gene fucosyltransferase 8 (FUT8) ($p=1.09 \times 10^{-8}$). This implies that fucosyltransferase 8 is the key regulator of glycan A2 concentration in human plasma. Allele A of this SNP alone explained 3.93% of the variance of glycan A2 (after adjustment for age and sex). This association is highly biologically plausible because glycan A2 is a known substrate for FUT8 (**18**). In addition to FUT8, there are three other enzymes that are directly involved in the metabolism of A2 glycan and which could potentially affect its plasma concentration (**18**): (i) α-1,6-mannosyl-glycoprotein 2-β-N-acetylglucosaminyltransferase, an enzyme that adds N-acetylglucosamine to mannoses and creates glycan A2; (ii) β-1,4-mannosyl-glycoprotein β1,4- N-acetylglucosaminyltransferase III, an enzyme that adds bisecting N-acetylglucosamine and decreases A2 by converting it to A2B; (iii) β-1,4-galactosyltransferase 1, an enzyme that adds galactose to N-acetylglucosamines and converts glycan A2 to A2G and A2G2 structures. We were not able to identify any SNPs in the genes encoding those enzymes that significantly affect glycan A2 concentrations in our study sample.

More detailed analyses provided further insights into the genetic regulation of glycan A2. In a genome-wide association study of glycan A2 levels prior to desialylation (sialidase digestion) of all glycans the peak on chromosome 14 was still apparent, but it barely reached genome-wide statistical significance. This implies that sialidase digestion improved the detectability of true genetic effects on a N-glycan and should be performed prior to HILIC HPLC analysis in genetic association studies. In addition, it was noted that the entire peak on Chr 14 can be explained by independent effects of only 2 SNPs: rs7161123 (in gene FUT8) and rs3020450, which also showed a strong association with glycan A2 ($p=6.56 \times 10^{-6}$) (**Table 1**). The latter SNP is located in a different gene (estrogen receptor beta - ESR2) and acts independently and additively with FUT8 to regulate levels of glycan A2. The independence between the action of two SNPs is illustrated by physical separation of the peaks on chromosome 14 on a higher resolution ($r^2 = 4.68 \times 10^{-04}$) (**Supplementary figure 1**) and the p-value for the effect of the interaction of the two SNPs on A2 glycan level was not statistically significant. Analysis of the combined effects of rs7161123 in FUT8 and rs3020450 in ESR2 on glycan A2 levels through multi-locus analysis showed p-value of 1.36×10^{-12} for the association of the haplotype with the trait in an additive model, with a combined effect of the two SNPs on the variance of glycan A2 (adjusted for age and sex) of 6.66%. This increased to 10.84% when the analysis was performed separately in females and decreased to 2.67% when performed only in males (**Supplementary table 1**).

Further analyses provided novel insights into the nature of genetic regulation of N-glycans in human plasma. A separate genome-wide association study by gender was performed, in which the values of glycan A2 were adjusted for age only. In this study conducted in 536 females rs7161123 in FUT8 was again the SNP most significantly associated with glycan A2 ($p=1.15 \times 10^{-9}$). An analysis in 388 males did not show any significant associations with glycan A2. The effect size of rs7161123 allele A in FUT8 in females was large, explaining 7.59% of age-adjusted trait variance (in comparison to 0.57% in males only). A similar study of rs3020450 in the ESR2 gene (A allele tested) showed the effect of SNP on the variance of the trait of 2.73% (3.25% in females and 2.10% in males). Thus, the association with FUT8 appears to be restricted to females, while ESR2 affects glycan A2 in both sexes.

The breakdown of females into premenopausal (N=166) and postmenopausal groups (N=352) and further association analysis showed that *FUT8* rs7161123 explains 17.31% of the trait variance (adjusted for age) in premenopausal women and 3.61% in postmenopausal women. In contrast, *ESR2* rs3020450 explained 0.51% of A2 glycan in premenopausal and 5.96% in postmenopausal women.

This is the first study of human N-glycan concentrations, measured as a quantitative trait, using a genome-wide association study. Recent technological advances in high-throughput glycome analyses combined with high-throughput genomics and mass spectrometry of proteins and lipids are set to reveal complex interactions between genes, proteins, lipids and glycans and unknown steps in metabolic pathways. This first insight into genetic regulation of plasma levels of an individual glycan (A2) showed strong sex-specific effects, differences in pre- and post-menopausal life and multi-locus effects, similar to many other quantitatively varying biological traits (**19,20**). It also showed that desialylation of glycan A2 before HPLC analysis reduced the measurement "noise" and so increased the power of the study to detect genetic effects. Although glycosylation is the most complex and abundant posttranslational modification, the importance of complex oligosaccharide structures attached to proteins for human health is only beginning to be appreciated (**17**). This is not surprising, since sequencing or synthesizing of branched sugar structures is significantly more challenging than the analysis of linear DNA and protein sequences (**11**). However, this study is encouraging because it shows that these difficulties can be overcome. In the next few years, we should expect new and exciting discoveries on the role of glycan structures in human biology, health and disease.

Supplementary Methods

Human samples. This study was based on samples from respondents who were residents of the Croatian island of Vis and who were recruited within a larger genetic epidemiology program which seeks to investigate genetic variability influencing common complex diseases and disease-related traits in genetically isolated populations (**21**). The island of Vis is one of the Dalmatian islands that has been moderately isolated throughout history. The sampling framework was based on the voting register which was used to send postal invitations to all adult inhabitants (over 18 years of age). The sample for this study consisted of a total of 1008 individuals aged 18-93 years (median age 56, interquartile range 24 years). There were 415 men (41.2%) and 593 women (58.8%). Genealogical records were reconstructed based on the Church Parish records and information provided by the respondents, and then checked against genetic data on allele sharing between relatives as a quality control measure to exclude data

errors. The sample contained a total of 809 genealogical relationships (including 205 parent-child, 123 sibling, and 481 other relationships). Blood samples were centrifuged at the time of collection and EDTA plasma aliquots stored at -70 °C until analysis. This study conformed to the ethical guidelines of the 1975 Declaration of Helsinki. All respondents signed informed consent before participating in the study and the study was approved by the appropriate Ethics Board of the University of Zagreb Medical School.

Glycan release and labeling. The N-glycans from plasma sample (5 µl) proteins were released and labeled with 2-aminobenzamide (LudgerTag 2-AB labeling kit Ludger Ltd., Abingdon, UK) as described previously (**22**). Labeled glycans were dried in a vacuum centrifuge and redissolved in known volume of water for further analysis.

Sialidase digestion. Aliquots of the 2-AB-labeled glycan pool were dried down in 200-µl microcentrifuge tubes. To these, the following was added: 1 µl of 500 mM sodium acetate incubation buffer (pH 5.5), 1 µl (0.005 units) of ABS, *Arthrobacter ureafaciens* sialidase (releases α2–3, 6, 8 sialic acid, Prozyme) and H_2O to make up to 10 µl. This was incubated overnight (16–18 h) at 37 °C and then passed through a Micropure-EZ enzyme remover (Millipore, Billerica, MA, USA) before applying to the HPLC.

Hydrophilic interaction high performance liquid chromatography (HILIC). Released glycans were subjected to hydrophilic interaction high performance liquid chromatography (HILIC) on a 250 x 4.6 mm i.d. 5µm particle packed TSKgel Amide 80 column (Anachem, Luton, UK) at 30°C with 50 mM formic acid adjusted to pH 4.4 with ammonia solution as solvent A and acetonitrile as solvent B. 60 min runs were on a 2795 Alliance separations module (Waters, Milford, MA). HPLCs were equipped with a Waters temperature control module and a Waters 2475 fluorescence detector set with excitation and emission wavelengths of 330 and 420 nm, respectively. The system was calibrated using an external standard of hydrolyzed and 2-AB-labeled glucose oligomers from which the retention times for the individual glycans were converted to glucose units (GU) (**23**). Glycans were analyzed on the basis of their elution positions and measured in glucose units then compared to reference values in the "GlycoBase" database (available at: http:// glycobase.nibrt.ie) for structure assignment (**24**).

.

Genotyping. A genome-wide association scan using 317,503 SNPs (Human Hap300, Illumina) was initially carried out in 991 Croatian individuals and scored using the BeadStudio software v.3 (Illumina). We removed a total of 7,096 SNPs that had <95% genotyping call rates, 1,289 that had a minor allele frequency <2% 113 that were out of Hardy-Weinburg equilibrium ($p<1 \times 10^{-10}$) and 9 X-linked markers that were likely to be autosomal, leaving 308,996 SNPs in the analyses.

Genome-wide association analysis. We removed 60 individuals who had low call rates (<97%), one individual with too high IBS (a known twin) and 6 individuals with sex discrepancies, leaving 924 individuals in this analysis. Genome-wide associations between glycan A2 phenotype and single nucleotide polymorphic (SNP) markers were analysed using the "mmscore" function of the GenABEL package for R statistical software (http://mga.bionet.nsc.ru/nlru/GenABEL), using an additive model. This score test for family based association (**25**) takes into account pedigree structure and allowed unbiased estimations of SNP allelic effect relatedness between examinees. The relationship matrix used in this analysis was generated by the "ibs" function of GenABEL which used IBS genotype sharing

to determine the realised pairwise kinship coefficient similarly to the PLINK "genome" function. All identified SNPs that reached significance or seemed to be suggestive of significance were visualised using HaploView software (http://www.broad.mit.edu/mpg/haploview/). The percentage of the variance explained by one or 2 SNPs was calculated using the "polygenic" function of the "GenABEL" package. Haplotype analysis was performed using the haplo.stat package for R (http://cran.r-project.org/web/packages/haplo.stats/index.html).

Author contributions

G.L. and I.R. designed the study and led the writing of the paper; H.C., A.F.W., P.M.R and N.D.H. designed the study and checked the manuscript for important intellectual content;., A.K. and O.G. performed laboratory analyses of glycans; O.P., I.K., L.Z. and Z.B. performed field work and constructed genealogies in Croatia; C.H., V.V. and J.H. carried out the genotyping and performed statistical analyses. All authors contributed to the preparation of the manuscript and declare that they have no conflicts of interest.

Acknowledgments

This work was supported by grants #219-0061194-2023 (to GL) and #108-1080315-0302 (to IR) from the Croatian Ministry of Science, Education and Sport; by European Commission FP6 EUROPHARM grant and EUROSPAN grant (number 018947 - LSHG-CT-2006-01947). The Vis study in the Croatian island of Vis was supported through the grants from the Medical Research Council UK to H.C., A.F.W. and I.R. The authors collectively thank to very large number of individuals for their individual help in organizing, planning and carrying out the field work related to the project and data management: Professor Pavao Rudan and the staff of the Institute for Anthropological Research in Zagreb, Croatia (organization of the field work, anthropometric and physiological measurements, and DNA extraction); Professor Ariana Vorko-Jovic and the staff and medical students of the Andrija Stampar School of Public Health of the Faculty of Medicine, University of Zagreb, Croatia (questionnaires, genealogial reconstruction and data entry); Dr Branka Salzer from the biochemistry lab "Salzer", Croatia (measurements of biochemical traits); local general practitioners and nurses (recruitment and communication with the study population); and the employees of several other Croatian institutions who participated in the field work, including but not limited to the University of Rijeka and Split, Croatia; Croatian Institute of Public Health; Institutes of Public Health in Split and Dubrovnik, Croatia.

Special thanks to K. Wilson and R. Bisset for administrative support.

References

1. Lauc G. Biochim. Biophys. Acta 2006; 1760:525-6.
2. Freeze HH. Nat. Rev. Genet. 2006; 7:537-51.
3. Spiro RG. Glycobiology 2002; 12:43R-56R.
4. Helenius A, Aebi M. Science 2001; 291:2364-9.
5. Marek KW, Vijay IK, Marth JD. Glycobiology 1999; 9:1263-71.
6. Jaeken J. J. Inherit. Metab. Dis. 2003; 26:99-118.

7. Vogt G, Chapgier A, Yang K, Chuzhanova N, Feinberg J, Fieschi C, et al. Nat. Genet. 2005; 37:692-700.
8. Vogt G, Vogt B, Chuzhanova N, Julenius K, Cooper DN, Casanova JL. Curr. Opin. Genet. Dev. 2007; 17:245-51.
9. Varki A. Glycobiology 1993; 3:97-130.
10. Gornik O, Royle L, Harvey DJ, Radcliffe CM, Saldova R, Dwek RA, Rudd P, Lauc, G. Glycobiology 2007; 17:1321-32.
11. Blow N. Nature 2009; 457:617-620.
12. Knezevic A, Polasek O, Gornik O, Rudan I, Campbell H, Hayward C, et al. J. Proteome Res. 2009.; (in press)
13. Vitart V, Rudan I, Hayward C, Gray NK, Floyd J, Palmer CN, et al. Nat. Genet. 2008; 40:437-42.
14. Rudd P, Rudan I, Wright AF. J Proteome Res 2009; (in press)
15. Altshuler D, Daly MJ, Lander ES. Science. 2008; 322:881-8.
16. Marth JD, Grewal PK. Nat. Rev. Immunol. 2008; 8:874-87.
17. Alavi A, Axford JS. Rheumatology (Oxford) 2008; 47:760-70.
18. Taniguchi N, Honke K, Fukuda M. Handbook of glycosyltransferases and related genes. Springer Verlag: Tokyo, 2002.
19. Ober C, Loisel DA, Gilad Y. Nat. Rev. Genet. 2008; 9:911-22.
20. Weiss LA, Pan L, Abney M, Ober C. Nat. Genet. 2006; 38:218-22.
21. Rudan I, Campbell H, Rudan P. Coll. Antropol. 1999; 23:531-46.
22. Royle L, Campbell MP, Radcliffe CM, White DM, Harvey DJ, Abrahams JL, et al. Anal. Biochem. 2008; 376:1-12.
23. Royle L, Radcliffe CM, Dwek RA, Rudd PM. Meth. Mol. Biol. 2006; 347:125-43.
24. Campbell MP, Royle L, Radcliffe CM, Dwek RA, Rudd PM. Bioinformatics 2008; 24:1214-6.
25. Chen WM, Abecasis GR. Am. J. Hum. Genet. 2007; 81:913-926.

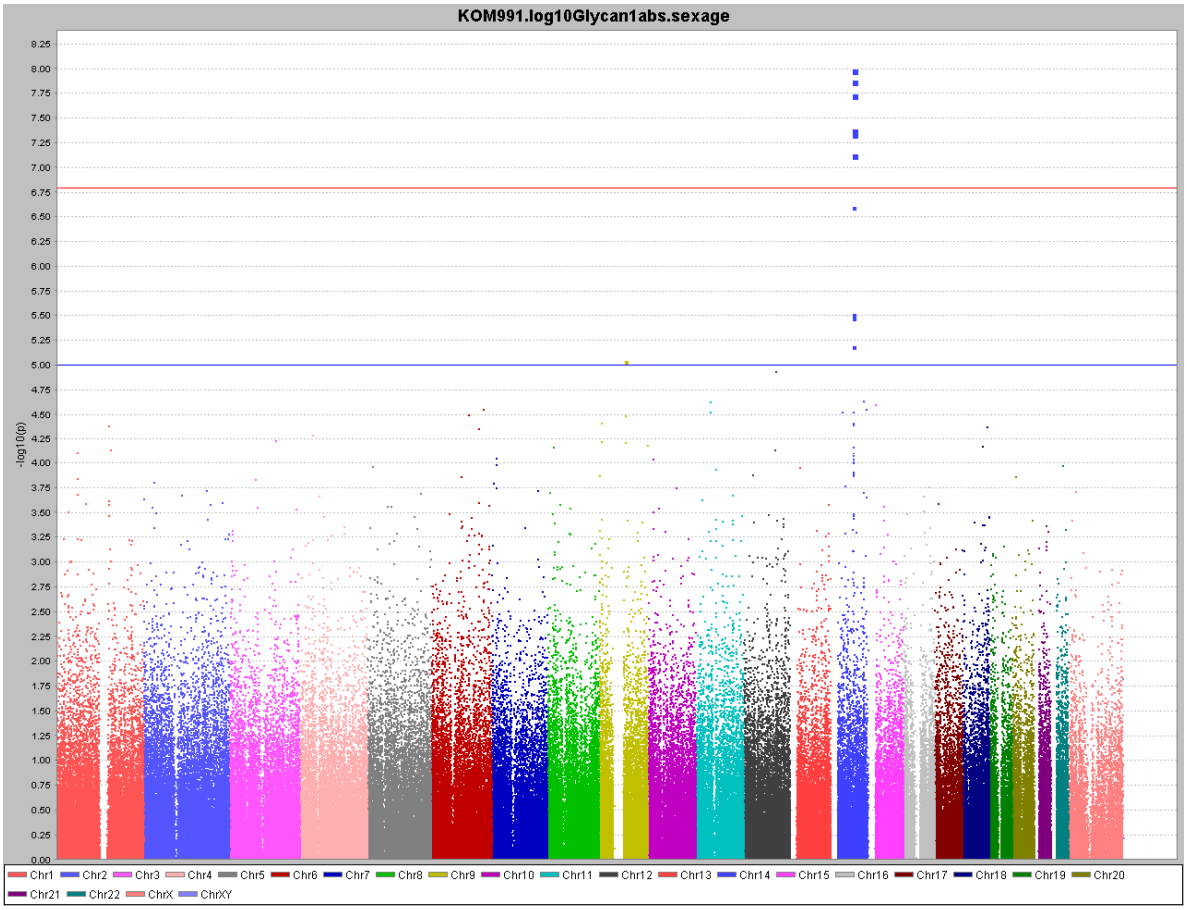
Table 1: A list of single nucleotide polymorphysms (SNPs) that showed association with glycan A2 reaching statistical significance of $p<10^{-5}$ in genome-wide association studies that included all examinees, females only and males only. Chromosome (chr) and position of each SNP, alleles found in studied sample, effective alleles (and their frequency in studied sample), size of the observed effect on glycan A2 (with 95% confidence intervals), p-value of significance of association with glycan A2 and implicated genes are shown for each SNP.

SNP	Chr	Position	Alleles	Effective Alleles	Eff. Allele Frequency	Effect Size	95% CI (Effect Size)		P-Value	Genes In Region
							Lower	Upper		
ALL EXAMINEES INCLUDED (N=924)										
rs6560442	9	77104027	G/A	A	0.0910	0.3180	0.1745	0.4615	9.31E-06	no genes +/- 100kb
rs8006145	14	63769203	C/A	A	0.2549	-0.1852	-0.2648	-0.1055	3.35E-06	ESR2
rs3020450	14	63838055	G/A	A	0.3072	-0.1539	-0.2223	-0.0856	6.56E-06	ESR2
rs7159888	14	64828395	G/A	A	0.4323	0.1243	0.0710	0.1777	3.10E-06	no genes +/- 100kb
rs11621121	14	64892246	A/G	G	0.4431	0.1380	0.0867	0.1893	7.44E-08	55kb 5' FUT8
rs2411822	14	64948148	A/G	G	0.4602	-0.1355	-0.1849	-0.0861	4.19E-08	FUT8
rs1953416	14	64948560	G/A	A	0.4566	-0.1289	-0.1788	-0.0789	2.52E-07	FUT8
rs7161123	14	65122654	G/A	A	0.4806	-0.1369	-0.1847	-0.0891	1.03E-08	FUT8
rs743085	14	65137886	A/G	G	0.4824	-0.1348	-0.1822	-0.0873	1.34E-08	FUT8
rs2411351	14	65241294	A/G	G	0.4851	-0.1329	-0.1802	-0.0857	1.84E-08	FUT8
rs10483785	14	65289270	G/A	A	0.4903	-0.1270	-0.1735	-0.0806	4.55E-08	10kb 3' FUT8
FEMALES ONLY INCLUDED (N=536)										
rs4852139	2	240475479	G/A	A	0.2758	-0.2150	-0.3120	-0.1180	9.34E-06	predicted gene region
rs11621121	14	64892246	A/G	G	0.4375	0.1886	0.1182	0.2590	8.31E-08	55kb 5' FUT8
rs8013442	14	64941614	G/A	A	0.3489	0.1895	0.1053	0.2737	6.80E-06	6kb 5' FUT8
rs2411822	14	64948148	A/G	G	0.4688	-0.1889	-0.2552	-0.1227	1.16E-08	FUT8
rs1953416	14	64948560	G/A	A	0.4686	-0.1790	-0.2449	-0.1131	5.57E-08	FUT8
rs1959144	14	65015804	C/A	A	0.3499	0.1905	0.1060	0.2750	6.50E-06	FUT8
rs4902399	14	65026650	G/A	A	0.3503	0.1918	0.1071	0.2764	5.85E-06	FUT8
rs3783711	14	65098199	G/A	A	0.3470	0.1873	0.1028	0.2718	9.24E-06	FUT8
rs7161123	14	65122654	G/A	A	0.4844	-0.1956	-0.2599	-0.1314	1.12E-09	FUT8
rs743085	14	65137886	A/G	G	0.4864	-0.1920	-0.2559	-0.1282	1.82E-09	FUT8
rs7144971	14	65190403	A/G	G	0.3426	0.1983	0.1120	0.2847	4.34E-06	FUT8
rs2411351	14	65241294	A/G	G	0.4892	-0.1825	-0.2456	-0.1194	7.30E-09	FUT8
rs10483785	14	65289270	G/A	A	0.4942	-0.1747	-0.2369	-0.1125	1.93E-08	10kb 3' FUT8
MALES ONLY INCLUDED (N=388)										
No SNPs reached genome-wide significance of $p<10^{-5}$										

Supplementary table 1: Analysis of the effects of haplotypes involving rs7161123 in FUT8 and rs3020450 in ESR2 in genome-wide association studies involving all examinees, males only and females only. P-values represent the statistical significance of association of haplotype with glycan A2 under additive model of the effects of two SNPs. Var. expl. denotes proportion of explained variance in measured levels of glycan A2 in the study sample.

Allele		Haplotype		
rs7161123	rs3020450	Frequency	Score	P-Value
All examinees (N=924; $p=1.356 \times 10^{-12}$; Var. expl.=6.66%)				
A	A	0.15228	-6.86666	6.57E-12
A	G	0.32830	-1.58025	0.1141
G	A	0.15489	-0.15728	0.8750
G	G	0.36452	6.46952	9.83E-11
Females only (N=536; $p=1.547 \times 10^{-11}$; Var. expl.=10.84%)				
A	A	0.14859	-6.1794	6.435E-10
A	G	0.33579	-2.62385	0.00869
G	A	0.15453	0.26854	0.78829
G	G	0.3611	6.4874	8.733E-11
Males only (N=388; p=0.00519; Var. expl.=2.67%)				
A	A	0.15736	-3.51292	0.00044
A	G	0.15548	-0.39137	0.69552
G	A	0.31790	0.53437	0.59308
G	G	0.36926	2.37867	0.01738

Figure 1. The result of genome-wide association study for N-linked A2 glycan (GlcNAc2Man3GlcNAc2) in human plasma of 986 individuals from Vis island, Croatia. HPLC analysis of fluorescently-labelled N-glycan A2 was followed by sialidase digestion and standardization by sex and age, which all increased detectability of underlying genetic variants. There is a clear "peak" at chromosome 14 including 6 single nucleotide polymorphysms showing association with A2 glycan beyond genome-wide significance level after Bonferroni correction ($P = 1.7 \times 10^{-7}$) represented by red horizontal line.



Supplementary figure 1. The independence between the action of rs7161123 in FUT8 and rs3020450 in ESR2 on glycan A2 levels is shown by physical separation of the peaks on chromosome 14 on a higher resolution. Correlation (r2) between the two SNPs is only $4.68x10^{-04}$. Analysis of the combined effects of rs7161123 in FUT8 and rs3020450 in ESR2 on glycan A2 levels shows strong haplotype effects and explains up to 10% of the trait variance.