Glycoconj J (2013) 30:41–50
DOI 10.1007/s10719-012-9397-y

Genomics and epigenomics of the human glycome

Vlatka Zoldoš · Mislav Novokmet · Ivona Bečeheli · Gordan Lauc

Received: 2 March 2012 / Revised: 11 May 2012 / Accepted: 14 May 2012 / Published online: 31 May 2012
© Springer Science+Business Media, LLC 2012

Abstract The majority of all proteins are glycosylated and glycans have numerous important structural, functional and regulatory roles in various physiological processes. While structure of the polypeptide part of a glycoprotein is defined by the sequence of nucleotides in the corresponding gene, structure of a glycan part results from dynamic interactions between hundreds of genes, their protein products and environmental factors. The composition of the glycome attached to an individual protein, or to a complex mixture of proteins, like human plasma, is stable within an individual, but very variable between individuals. This variability stems from numerous common genetic polymorphisms reflecting in changes in the complex biosynthetic pathway of glycans, but also from the interaction with the environment. Environment can affect glycan biosynthesis at the level of substrate availability, regulation of enzyme activity and/or hormonal signals, but also through gene-environment interactions. Epigenetics provides a molecular basis how the environment can modify phenotype of an individual. The epigenetic information (DNA methylation pattern and histone code) is especially vulnerable to environmental effects in the early intrauterine and neo-natal development and many common late-onset diseases take root already at that time. The evidences showing the link between epigenetics and glycosylation are accumulating. Recent progress in high-throughput glycomics, genomics and epigenomics enabled first epidemiological and genome-wide association studies of the glycome, which are presented in this mini-review.

Keywords Glycosylation · Glycome · Genome-wide association study · Epigenetics · Gene-environment interactions

V. Zoldoš (✉)
University of Zagreb, Faculty of Science,
Horvatovac 102a,
Zagreb, Croatia
e-mail: vzoldos@biol.pmf.hr

M. Novokmet · I. Bečeheli · G. Lauc
Genos Ltd, Glycobiology Laboratory,
Planinska 1,
Zagreb, Croatia

G. Lauc (✉)
University of Zagreb, Faculty of Pharmacy and Biochemistry,
A. Kovačića 1,
Zagreb, Croatia
e-mail: glauc@pharma.hr

Genetics of protein glycosylation is very complex

According to the central dogma of molecular biology, function of each protein is determined by its structure, which is defined by the nucleotide sequence in the corresponding gene. However, in the case of glycan moieties of glycoproteins, there are several additional layers of complexity between genes and the final glycan structure. The final structure of each glycan is therefore not encoded directly in the genome (with one or two genes) but many genes determine the structure of proteins (enzymes and other proteins) that participate in the complex biosynthetic pathway of glycan synthesis. Genetic polymorphisms in the participating genes, regulation of gene expression, posttranslational modifications, and the activity of the corresponding proteins all work together to determine final structure of a glycan. Additional mechanisms, including altered intracellular localization, competition with endogenous acceptor substrates and variable access to monosaccharide donor substrates can also affect the final outcome [1].

Glycan structures are generated by concerted action of hundreds of glycosidases and glycosyltransferases [2],

which are probably only a tip of the iceberg of the proteins involved in the glycosylation process. Recent study in *Drosophila* revealed that only 13 % out of 109 genes found to be needed for neural-specific glycosylation were previously known glyco-genes [3]. However, the remaining 87 % of genes needed for neural-specific glycosylation probably also participate in other types of cellular processes and cannot be claimed to be new "glyco-genes". Also the same genes are probably needed for other types of glycosylation not covered in this study. Therefore direct extrapolation of this ratio is not possible, but it is highly probable that at least two or three times more genes than currently listed 600 glyco-genes are required for protein glycosylation. Therefore, at least 5 % of the human genome seems to be involved in protein glycosylation, suggesting that this biological process is the most complex biosynthetic pathway in multicellular organisms.

Since protein structure is defined by the sequence of nucleotides in the corresponding gene, genetic polymorphisms affect protein structure in a known and mostly predictable way. With glycans this is frequently not the case because polymorphisms in different enzymes and other proteins involved in protein glycosylation pathways can interact in a very complex manner. For example, a polymorphism in a given glycosyltransferase could lead to 10 % decrease in the activity of this enzyme due to lower affinity of an enzyme for the activated sugar nucleotide. Consequently, glycans will bear slightly lower amount of this specific monosaccharide and the composition of the entire glycome will be altered. However, if the same individual also has a polymorphism in other genes in the same biosynthetic pathway or in the transmembrane transporter of the activated sugar nucleotide (which decreases its concentration in the Golgi), the combined effect of decreased enzyme activity and decreased substrate concentration could be very prominent. Alternatively, the same individual can have a polymorphism in a proton transporter and altered pH in Golgi, which could actually increase the activity of that specific glycosyltransferase (or decrease activity of a competing glycosyltransferase) and effectively erase the effect of the original polymorphism on the glycome composition.

Therefore, complex epistatic interactions among multiple gene loci determine the final glycan structure. This is nicely illustrated by AB0 system of blood group determination, which arises from the existence of the three allelic variants (A, B, and 0) of a single glycosyltransferase gene at the same locus (AB0 locus); allele A is coding for *N*-acetyl-galactosaminyltransferase, allele B is coding for galactosyltransferase and allele 0 is a non-functional truncated variant [4]. The two enzymes add different sugars to a five-monosaccharide glycan precursor (H-antigen); *N*-acetyl-galactosaminyltransferase adds *N*-acetylgalactosamine and galactosytransferase adds galactose to H-antigen to produce A- or B- antigens, respectively. Another enzyme, fucosyltransferase is responsible for the production of the H antigen, and this enzyme is encoded by a gene at different locus (H). In the human population, there are approximately 15 % of recessive homozygotes for this gene (hh genotype) that cannot produce the H antigen (Bombay phenotype). Therefore, recessive homozygotes for the gene H will appear as type 0 blood group regardless of their genotype at the AB0 locus. Two persons, each carrier of one of the two different mutations (the carrier of the recessive allele h and the carrier of the allele 0) will show the same external phenotype following the blood test—there will be no coagulation with antibodies against A or B antigens in both persons, even if their biochemical phenotypes are different (the absence of H antigen in carriers of mutation in H gene, and the presence of H-antigen in carriers of deletion in AB0 gene). This example of the interaction between only two different glyco-genes illustrates a complexity in production of glycoprotein structures and its repercussions on the glycome composition and eventually on the biochemical and external phenotype.

Genome-wide association studies initiated a new revolution in genetics

Until about 5 years ago two major designs were used to study the role of human genetic variation in common complex diseases: genetic linkage studies and candidate-gene approaches based on *a priori* hypotheses of biological plausibility. In the studies of genetic linkage scarce set of genetic markers were usually used and they have only been successful in identifying rare genetic mutations with high penetrance and large effects in extended pedigrees affected by monogenic ("Mendelian") disorders. These studies were generally underpowered to detect more common genetic variants of smaller effect sizes in the samples drafted from the general population [5]. On the other hand, the majority of candidate gene studies used sample sizes of up to several hundred cases and controls, what led to results that were generally not replicated in subsequent studies [6, 7].

However, a major technological breakthrough occurred when HapMap project outcome was combined with the advances in genotyping technologies. In the course of the HapMap project, more than 3 million single nucleotide polymorphysms (SNPs) were identified. Taking into account recombination hotspots in the genome and their linkage disequilibrium (LD) patterns, the selection of several hundred thousand of particularly useful SNP markers (haplotype tagging) was made. This has resulted in the development of high density SNP arrays with hundreds of thousands of SNP markers spanning across the entire genome, which made hypothesis-free genome-wide association studies (GWAS) possible [8].

Further on, issues such as population stratification, relatedness of the subjects in the sample and multiple testing had to be dealt with. To address those problems new advanced, statistical and computational methods were developed with agreement on a very stringent genome wide statistical significance level of approximately $p=5\times10^{-7}$ when a set of about 300,000 SNPs is used. By accepting those criteria, false-positive results that plagued the field for more than a decade were resolved [9, 10]. The studies were initially conducted in reasonably small sample sizes of up to 2,000 cases and controls. Initial successes soon led to formation of very large consortia and pooling of samples among several studies in genome wide association study meta-analyses. At the moment large consortia such as GIANT (which studies anthropometric traits) or DIAGRAM (which studies type 2 diabetes) are incorporating about a hundred thousand of cases and controls. Hundreds of bio-medically relevant quantitative traits and common genetic variants underlying complex human diseases have been identified and widely replicated across populations. Genetic networks influencing a wide range of common complex diseases have begun to emerge, including type 1 and 2 diabetes, ischaemic heart disease, stroke, rheumatoid arthritis, breast/prostate/colorectal/lung cancer, and inflammatory bowel disease; and a range of endophenotype underlying these diseases such as obesity, LDL cholesterol and fasting glucose. Emerged variants typically showed small effect sizes (odd ratios of 1.1 to 1.4) and many have been located in regions outside the known genes, presumably in distant regulatory regions. They individually only explain a very small percentage of the overall variance in disease or traits and thus cannot be used, at this point in time, for a reliable genetic profiling of disease risk in individual subjects. However, rather large number of new aetiological hypotheses has been generated from those studies. A great number of GWAS accomplished in the last few years have formed a robust base of new knowledge from which functional studies can now rise to explore new disease pathways [11–13]. All of this should eventually result in the improved prediction of disease risk and will, with new class of drugs developed due to knowledge generated by these studies, contribute to the advance in highly personalized era of medicine.

Genes and environment integrate in the process of glycan biosynthesis

An important source of complexity and variability of the glycome is the interaction with the environment. Biosynthesis of glycans requires many hexosamine building blocks and their availability significantly affects structure of glycans and composition of the glycome [14]. Altered pH in Golgi [15, 16], oxygen concentration [17] and many other external factors also affect protein glycosylation. The expression level and activity of all enzymes in the complex pathway of glycan biosynthesis is also not pre-defined, but changes in response to various internal and external factors.

It is well established that the epigenome is a key mediator of gene-environment interactions [18]. While environmental factors, such as diet or chemical compounds, influence enzymatic processes only while they are directly present, their prolonged effects can be achieved through the epigenetic cell memory. The epigenomic program (DNA methylation pattern and histone code) is being established during early stages of intrauterine and neonatal development, and is particularly sensitive to environmental influences (dietary, chemical, behavioural, etc.) in that period [19]. For example, specific diet (such as exposure to famine, or intake of food rich in folate or betain) or exposure to certain environmental factors (such as cigarette smoke, aflatoxin B1 or UV radiation) in early life was reported to be critical for the establishment of the DNA methylation pattern with life-long effects for the individual [20, 21]. The epigenetic principle explains how environmental exposures can have long-lasting influence on biology and health [22, 23] without changing the gene sequence.

In response to environmental and extrinsic stimuli, epigenetic mechanisms (DNA methylation, RNA-mediated silencing and chromatin modifications) regulate gene expression [18] and the alteration in the genome expression program is inherited through cell divisions. Thus, the mitotic stability of epigenetic information is responsible for the link of the early experience with the future phenotype outcome (health). Indeed, the vast majority of late-onset common diseases often have epigenetic roots, mostly during early individual development [24, 25]. This is well established for many cancers, most notably for lung cancer [24], diabetes [26] and other multifactorial diseases. In addition, significant environmentally induced stochastic epigenetic changes occur during lifetime of an individual and increase the risk of age-associated diseases, but it is difficult to unravel the relative contribution of genetic (mutations and polymorphisms) and epigenetic alterations in a changed phenotype.

The evidences showing the link between epigenetics and glycosylation are accumulating, but, until now, they mostly reported aberrant glycosylation in cancer. It has been shown that a change in cytosine methylation within promoter of certain glyco-genes is responsible for the expression of cancer-associated carbohydrate antigens in gastrointestinal, colon, pancreatic, and breast cancer [27–30]. Other examples of epigenetic regulation of glyco-genes include: *FUT7* in leukocytes [27], *FUT4*, GDP-fucose transporter and *FX* genes involved in regulation of global fucosylation [31], and the transcription factor HNF1A, a master regulator of plasma protein fucosylation [32]. Early studies on DNA methylation

and expression levels of *HNF1A* in hepatocellular carcinomas (HCC) different in origin (HBV, HCV and alcohol induced HCC) showed that tumors induced by chronic alcohol intake differ from other two types of HCCs (manuscript in preparation). This finding indicates that environmental factors have strong influence on epigenetic regulation of the gene involved in glycosylation. Revolutionary finding of meiotically stable epialleles in humans has recently emerged. Even still scarce, there are evidences of epialleles passing to offspring and even grand-offspring through germline cells, putting the transgenerational epigenetic inheritance [33–35] in the focus of the epigenetic community. The importance of this phenomenon lies in the fact that environment and the individual experiences can shape biology across the lifecycle and even within lineages through time. It is tempting to speculate that epigenetic regulation of glycosylation is an essential evolutionary mechanism, which enables quick adaptation of complex organisms to environmental changes and help them avoid rapidly evolving pathogenic microorganisms.

Glycome variability in a human population

The combined effects of all genetic, epigenetic and environmental factors are reflected in high variability of glycome composition in a human population. For example, the smallest N-glycan in human plasma (A2, bi-antennary glycan with structure $GlcNAc_2Man_3GlcNAc_2$) was reported to represent between 0.03 % and 0.83 % of the total plasma N-glycome [36]. Median value for this glycan was 0.16, thus both the lowest and the highest observed level were very different from "normal" values. GWAS of this glycan trait identified a number of SNPs in *FUT8* gene, which explained between 1 % and 6 % of the variance [37, 38]. Heritability for this glycan was estimated to be approximately 50 % [36], thus the majority of heritable polymorphisms that affect this glycan are still not identified. The same is true for environmental factors; all measured variables like age, gender, BMI, body mass, biochemical parameters, and/or smoking, etc., cumulatively explained only 5 % of the variance [39], thus main non-heritable factors, which affect synthesis of this glycan remain to be identified.

The variability of the composition of the plasma glycome derives from both variability in the composition of the plasma proteome and the variability in the glycosylation process. However, when glycome composition of the isolated IgG was analysed, the variability was even greater than in the total plasma glycome. A2 glycan was found to represent between 0.1 % and 9.4 % of the IgG glycome, what is nearly a 100-fold difference [40] (the median value was 0.7 %). The average ratio between maximal and minimal observed values for all glycans in the IgG glycome was 19.7, what is nearly three times higher than the analogous ratio in the plasma glycome [36, 40]. Therefore, assuming that variability in glycosylation of other proteins will be similar to the variability observed in IgG, it appears that the large variability observed in the composition of the plasma glycome might actually be an averaged sum of even larger variability in the composition of glycomes of individual proteins present in the human plasma.

This may not be a universal rule, but for some proteins at least, this variability in glycosylation has significant functional consequences. For example, effector functions of IgG are largely regulated by glycosylation [41]. The activity of various membrane receptors and transporters is also by large regulated by their glycosylation [42, 43]. Recently, it was reported that in a large set of experimentally determined mouse N-glycosylation sites, the evolutionary rate of glycosylated asparagines was significantly lower than that of non-glycosylated asparagines of the same proteins [44]. This widespread and strong functional constraint on glycosylation, unlike what has been observed for phosphorylation sites, implies an overall functional importance of N-glycosylation.

Genome wide association studies of the human glycome

Three genome wide association studies (GWAS) of glycosylation-related traits were published until now [37, 38, 45]. The first GWAS of the glycome was performed on 2.705 individuals from Croatian islands Vis and Korčula and Orkney islands in Scotland. The composition of the desialylated total plasma glycome was determined by HPLC analysis that separated the total glycome into 13 HPLC peaks. Two further traits were derived from the original variables to calculate the percentage of glycan structures containing core (FUC-C) or antennary (FUC-A) fucose. Genome-wide significant associations were found for five HPLC peaks, as well as FUC-A [37]. Two of the identified genes were known glycosyltransferase genes, *FUT6* and *FUT8*. They associated with glycan structures that are known substrates or products of the corresponding glycosyltransferases, thus molecular mechanisms behind these associations were clear. The third identified gene was a transcription factor HNF1A, which was previously not known to affect protein glycosylation. The subsequent RNAi and ChIP studies have clearly demonstrated that HNF1A acts as a master regulator of plasma protein antennary fucosylation by promoting both *de novo* and the salvage pathways of GDP-fucose synthesis, stimulating the expression of *FUT3*, *FUT5* and *FUT6*, and supressing the expression of *FUT8* [37].

The second published GWAS of the glycome was extension of the first study, which both encompassed more individuals (3,533) and more detailed glycome analysis [38]. All findings from the first study were confirmed. This study revealed that *HNF1A* associates not only with fucosylation but also with glycan branching. Three new genes found to associate with

plasma glycome composition were *MGAT5, B3GAT1* and *SLC9A9*. *MGAT5* was found to associate with the proportion of highly branched glycans, what is in accordance with the known biological function of this gene. The second identified gene, *B3GAT1*, is a member of the glucuronyltransferase gene family. Product of this gene adds glucuronic acid to the terminal *N*-acetyllactosamine (Lac) disaccharide to form the HNK-1 epitope precursor [46, 47]. The HNK-1 epitope is expressed on a subset of human lymphocytes, but it was not previously reported to exist on plasma proteins. To explain the association of *B3GAT1* with the composition of the plasma glycome more detailed structural analysis was performed, which confirmed the existence of glucuronic acid on a subset of N-glycans released from human plasma glycoproteins. The third identified gene, *SLC9A9*, codes for a proton pump, which affects pH in the endosomal compartment [48]. This gene was not previously linked to glycosylation, but it is known that changes in Golgi pH can impair protein sialylation [16], thus the association between *SLC9A9* and tetrasialylated glycans makes biological sense.

Recent large RNAi study identified 109 genes, which are needed for neural-specific glycosylation in *Drosophila* [3] only 14 among them were previously known glyco-genes, while others were involved in various other cellular functions. Two GWAS studies described above identified four glycosyltransferases and two genes that were previously not linked to glycosylation. The difference in the ratio of glyco-genes and other genes identified using these two approaches is striking, and it may be explained by the fact that RNAi library approach has focused on neural specific glycosylation, while GWAS studies analysed composition of the total plasma glycome. Plasma glycans originate from different cell types and since glycosylation is known to be cell type-specific, the regulation of glycosylation must also be cell type-specific. By analysing pooled glycans from different proteins from different cells, these cell type-specific effects are blurred. Therefore, GWAS of glycomes of individual proteins will be needed to identify genes that regulate cell type-specific glycosylation.

Glycosyltransferases in genome wide association studies

Thousands of GWAS studies conducted between 2007 and 2012 identified over 3,000 SNPs, which were associated with various traits with genome-wide significance. Many of these polymorphisms were located within or near to known glycosyltransferase genes (Table 1). Allelic variants in glycosyltransferases were found to associate with basic biochemical parameters like cholesterol and triglycerides, enzymatic activity (alkaline phosphatase, gamma - glutamyltransferase), and even some general human traits like height and obesity. Molecular mechanisms behind these associations are generally not known. For the association between polymorphisms in glycosyltransferases and the activity of plasma enzymes it could be speculated that proper glycosylation is needed for their secretion or enzymatic activity, but for quantitative traits like red blood count, obesity, or height, any speculation about the molecular mechanism would be completely unjustified.

Allelic variants in glycosyltransferase genes were also found to associate with some diseases (Table 1). For example, a comprehensive study of genes that mediate breast cancer metastasis to the brain identified *ST6GALNAC5* (α2,6-sialyltransferase) as the key gene, which specifically mediates brain metastasis [49]. Normally restricted to the brain, the expression of *ST6GALNAC5* in breast cancer cells enhances their adhesion to brain endothelial cells and their passage through the blood–brain barrier. This co-option of a brain sialyltransferase highlights the role of the cell-surface glycosylation in organ-specific metastatic interactions. Another example is multiple sclerosis where *MGAT5* was reported as the most significant top hit in a recent GWAS for variants regulating disease severity [50]. *MGAT5* is known to affect membrane half-life of numerous proteins [42, 51], thus its association with disease is not unexpected, but exact molecular mechanism still has to be revealed.

Future prospects of genomics and epigenomics of glycosylation

The majority of all proteins are glycosylated. For some of them the exact structure of a glycan part is perhaps not important and a large number of different glycans could perform the same task equally well. However, on many glycoproteins, a glycan part has important structural, functional and regulatory role and the alternative glycosylation is important in many physiological processes [41]. Consequently, differences in glycosylation are probably important element in development and progression of virtually all complex diseases. Numerous glyco-biomarkers have been identified [52–55], but molecular mechanisms linking differences in glycosylation and disease pathology are generally not known. This is not surprising, since very little is known about the role of glycans on individual glycoproteins. Perhaps the only exception is IgG where the importance of different glycoforms for effector functions has been clearly demonstrated (for a review see [41]).

In general, two types of effects can be expected from common polymorphisms in glycosylation: (i) general effects and (ii) effects confined to a specific glycoprotein, a subset of glycoproteins or a tissue. A polymorphism in a gene will result in less efficient (or less expressed) specific glycosyltransferase and consequently all glycoproteins, which are substrates for that glycosyltransferase, will be affected. At the glycome

Table 1 Published genome-wide significant associations of diseases or quantitative traits with genetic polymorphisms in glycosyltransferases

Gene	Gene name	Disease (trait)	Ref.
ABO (A3GALNT, A3GALT1)	Transferase A, alpha 1-3-N-acetylgalactosaminyltransferase; Transferase B, alpha 1-3-galactosyltransferase	Alkaline phosphatase	[57–59]
		Angiotensin-converting enzyme activity	[60]
		Hematocrit	[57]
		Hemoglobin concentration	[57]
		Red blood cell count	[57]
		Serum phytosterol	[61]
		Serum soluble E-selectin	[62, 63]
		Serum soluble P-Selectin	[64]
		Soluble ICAM-1	[64, 65]
		TNF-α levels	[66]
		Venous thromboembolism	[67]
MGAT1	Mannosyl (α-1,3-)-glycoprotein β-1,2-N-acetylglucosaminyltransferase	Obesity	[68]
MGAT5	Mannosyl (α-1,6-)-glycoprotein β-1,6-N- acetylglucosaminyltransferase	Multiple sclerosis	[50]
ST6GAL1	CMP-N-acetylneuraminate-β-galactosamine-α-2, 6-sialyltransferase 1	Drug-induced liver injury (flucloxacillin)	[69]
ST3GAL4	β-galactoside α-2,3-sialyltransferase 4	Alkaline phosphatase	[59]
ST6GALNAC5	α2,6-sialyltransferase	Cancer metastasis to the brain	[49]
FUT2	Galactoside 2-alpha-L-fucosyltransferase 2	Crohn's disease	[70, 71]
		Vitamin B12	[72, 73]
		Gama - glutamyltransferase	[59]
		Alkaline phosphatase	[59]
FUT8	α-(1,6)-fucosyltransferase	N-glycan levels	[37]
GALNT2	Polypeptide N-acetylgalactosaminyltransferase 2	Cholesterol, HDL	[74–76]
		Triglycerides	[75]
B3GNT2	UDP-GlcNAc:βGal β-1,3-N-acetylglucosaminyltransferase 1	Ankylosing spondylitis	[77]
GLT6D1	Glycosyltransferase 6 domain-containing protein 1	Periodontitis	[78]
GLT25D2	Glycosyltransferase 25 family member 2	Height	[79]

level, there will be a change in the composition, which will reflect the decreased activity of the affected glycosyltransferase. Since various proteins will be affected, this pleiotropic effect will also vary significantly between individuals. Alternatively, a polymorphism might have an effect only on a specific glycoprotein expressed in a specific tissue. In that case, the consequences can be very specific because they would reflect decreased activity of a single enzyme, altered receptor specificity or intracellular localization. An interesting example is the miss-localisation of GLUT receptors in pancreatic islets of mice fed with high-fat diet [56]. Specific genetic polymorphisms or epigenetic silencing of a specific gene might have similar consequences and consequently make an individual more prone to diabetes.

Genomic studies of the human glycome are still very scarce, but proof of principle studies of association between glycome composition and both genetic polymorphisms [37, 38] and epigenetic marks [32] have recently been published. The main obstacle is the need to reliably quantify glycans in thousands of individuals, but with progress in high-throughput technology this is becoming feasible. Further genomic and epigenomic studies, which will contribute to the understanding of the role of glycosylation in both normal physiology and disease, are expected soon.

Acknowledgments The work in author's laboratory is supported by the Croatian Ministry of Science, Education and Sport grant #309-0061194-2023 (to GL) and #119-1191196-1224 (to VZ), and by the European Commission GlycoBioM (contract #259869) and HighGlycan (contract #278535) grants.

References

1. Ohtsubo, K., Marth, J.D.: Glycosylation in cellular mechanisms of health and disease. Cell **126**(5), 855–867 (2006)
2. Nairn, A.V., York, W.S., Harris, K., Hall, E.M., Pierce, J.M., Moremen, K.W.: Regulation of glycan structures in animal tissues: transcript profiling of glycan-related genes. J. Biol. Chem. **283**(25), 17298–17313 (2008)

3. Yamamoto-Hino, M., Kanie, Y., Awano, W., Aoki-Kinoshita, K.F., Yano, H., Nishihara, S., Okano, H., Ueda, R., Kanie, O., Goto, S.: Identification of genes required for neural-specific glycosylation using functional genomics. PLoS Genet. **6**(12), e1001254 (2010)
4. Yoshida, A.: Identification of genotype of blood group A and B. Blood **55**(1), 119–123 (1980)
5. Wright, A., Charlesworth, B., Rudan, I., Carothers, A., Campbell, H.: A polygenic basis for late-onset disease. Trends Genet. **19**(2), 97–106 (2003)
6. Chanock, S.J., Manolio, T., Boehnke, M., Boerwinkle, E., Hunter, D.J., Thomas, G., Hirschhorn, J.N., Abecasis, G., Altshuler, D., Bailey-Wilson, J.E., Brooks, L.D., Cardon, L.R., Daly, M., Donnelly, P., Fraumeni Jr., J.F., Freimer, N.B., Gerhard, D.S., Gunter, C., Guttmacher, A.E., Guyer, M.S., Harris, E.L., Hoh, J., Hoover, R., Kong, C.A., Merikangas, K.R., Morton, C.C., Palmer, L.J., Phimister, E.G., Rice, J.P., Roberts, J., Rotimi, C., Tucker, M.A., Vogan, K.J., Wacholder, S., Wijsman, E.M., Winn, D.M., Collins, F.S.: Replicating genotype-phenotype associations. Nature **447**(7145), 655–660 (2007)
7. Ioannidis, J.P.: Why most published research findings are false. PLoS Med. **2**(8), e124 (2005)
8. Cardon, L.R., Bell, J.I.: Association study designs for complex diseases. Nat. Rev. Genet. **2**(2), 91–99 (2001)
9. Vitart, V., Rudan, I., Hayward, C., Gray, N.K., Floyd, J., Palmer, C.N., Knott, S.A., Kolcic, I., Polasek, O., Graessler, J., Wilson, J.F., Marinaki, A., Riches, P.L., Shu, X., Janicijevic, B., Smolej-Narancic, N., Gorgoni, B., Morgan, J., Campbell, S., Biloglav, Z., Barac-Lauc, L., Pericic, M., Klaric, I.M., Zgaga, L., Skaric-Juric, T., Wild, S.H., Richardson, W.A., Hohenstein, P., Kimber, C.H., Tenesa, A., Donnelly, L.A., Fairbanks, L.D., Aringer, M., McKeigue, P.M., Ralston, S.H., Morris, A.D., Rudan, P., Hastie, N.D., Campbell, H., Wright, A.F.: SLC2A9 is a newly identified urate transporter influencing serum urate concentration, urate excretion and gout. Nat. Genet. **40**(4), 437–442 (2008)
10. Aulchenko, Y.S., Ripatti, S., Lindqvist, I., Boomsma, D., Heid, I.M., Pramstaller, P.P., Penninx, B.W., Janssens, A.C., Wilson, J.F., Spector, T., Martin, N.G., Pedersen, N.L., Kyvik, K.O., Kaprio, J., Hofman, A., Freimer, N.B., Jarvelin, M.R., Gyllensten, U., Campbell, H., Rudan, I., Johansson, A., Marroni, F., Hayward, C., Vitart, V., Jonasson, I., Pattaro, C., Wright, A., Hastie, N., Pichler, I., Hicks, A.A., Falchi, M., Willemsen, G., Hottenga, J.J., de Geus, E.J., Montgomery, G.W., Whitfield, J., Magnusson, P., Saharinen, J., Perola, M., Silander, K., Isaacs, A., Sijbrands, E.J., Uitterlinden, A.G., Witteman, J.C., Oostra, B.A., Elliott, P., Ruokonen, A., Sabatti, C., Gieger, C., Meitinger, T., Kronenberg, F., Doring, A., Wichmann, H.E., Smit, J.H., McCarthy, M.I., van Duijn, C.M., Peltonen, L.: Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts. Nat. Genet. **41**(1), 47–55 (2009)
11. Rudan, I., Marusic, A., Jankovic, S., Rotim, K., Boban, M., Lauc, G., Grkovic, I., Dogas, Z., Zemunik, T., Vatavuk, Z., Bencic, G., Rudan, D., Mulic, R., Krzelj, V., Terzic, J., Stojanovic, D., Puntaric, D., Bilic, E., Ropac, D., Vorko-Jovic, A., Znaor, A., Stevanovic, R., Biloglav, Z., Polasek, O.: 10001 Dalmatians: Croatia launches its national biobank. Croat. Med. J. **50**(1), 4–6 (2009)
12. Hirschhorn, J.N.: Genomewide association studies–illuminating biologic pathways. N. Engl. J. Med. **360**(17), 1699–1701 (2009)
13. Goldstein, D.B.: Common genetic variation and human traits. N. Engl. J. Med. **360**(17), 1696–1698 (2009)
14. Lau, K.S., Partridge, E.A., Grigorian, A., Silvescu, C.I., Reinhold, V.N., Demetriou, M., Dennis, J.W.: Complex N-glycan number and degree of branching cooperate to regulate cell proliferation and differentiation. Cell **129**(1), 123–134 (2007)
15. Hassinen, A., Pujol, F.M., Kokkonen, N., Pieters, C., Kihlstrom, M., Korhonen, K., Kellokumpu, S.: Functional organization of golgi N- and O-Glycosylation pathways involves ph-dependent complex formation that is impaired in cancer cells. J. Biol. Chem. **286(44), 38329–38340 (2011)**
16. Rivinoja, A., Hassinen, A., Kokkonen, N., Kauppila, A., Kellokumpu, S.: Elevated Golgi pH impairs terminal N-glycosylation by inducing mislocalization of Golgi glycosyltransferases. J. Cell. Physiol. **220**(1), 144–154 (2009)
17. Kunkel, J.P., Jan, D.C., Jamieson, J.C., Butler, M.: Dissolved oxygen concentration in serum-free continuous culture affects N-linked glycosylation of a monoclonal antibody. J. Biotechnol. **62** (1), 55–71 (1998)
18. Jaenisch, R., Bird, A.: Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. Nat. Genet. **33**(Suppl), 245–254 (2003)
19. Faulk, C., Dolinoy, D.C.: Timing is everything: the when and how of environmentally induced changes in the epigenome of animals. Epigenetics **6**(7), 791–797 (2011)
20. Heijmans, B.T., Boer, J.M., Suchiman, H.E., Cornelisse, C.J., Westendorp, R.G., Kromhout, D., Feskens, E.J., Slagboom, P.E.: A common variant of the methylenetetrahydrofolatereductase gene (1p36) is associated with an increased risk of cancer. Cancer Res. **63**(6), 1249–1253 (2003)
21. Heijmans, B.T., Tobi, E.W., Stein, A.D., Putter, H., Blauw, G.J., Susser, E.S., Slagboom, P.E., Lumey, L.H.: Persistent epigenetic differences associated with prenatal exposure to famine in humans. Proc. Natl. Acad. Sci. U.S.A. **105**(44), 17046–17049 (2008)
22. Barker, D.J., Eriksson, J.G., Forsen, T., Osmond, C.: Fetal origins of adult disease: strength of effects and biological basis. Int. J. Epidemiol. **31**(6), 1235–1239 (2002)
23. Tobi, E.W., Lumey, L.H., Talens, R.P., Kremer, D., Putter, H., Stein, A.D., Slagboom, P.E., Heijmans, B.T.: DNA methylation differences after exposure to prenatal famine are common and timing- and sex-specific. Hum. Mol. Genet. **18**(21), 4046–4053 (2009)
24. Herceg, Z., Vaissiere, T.: Epigenetic mechanisms and cancer: an interface between the environment and the genome. Epigenetics **6** (7), 804–819 (2011)
25. Thayer, Z.M., Kuzawa, C.W.: Biological memories of past environments: epigenetic pathways to health disparities. Epigenetics **6** (7), 798–803 (2011)
26. Ling, C., Groop, L.: Epigenetics: a molecular link between environmental factors and type 2 diabetes. Diabetes **58**(12), 2718–2725 (2009)
27. Syrbe, U., Jennrich, S., Schottelius, A., Richter, A., Radbruch, A., Hamann, A.: Differential regulation of P-selectin ligand expression in naive versus memory CD4+ T cells: evidence for epigenetic regulation of involved glycosyltransferase genes. Blood **104**(10), 3243–3248 (2004)
28. Kawamura, Y.I., Toyota, M., Kawashima, R., Hagiwara, T., Suzuki, H., Imai, K., Shinomura, Y., Tokino, T., Kannagi, R., Dohi, T.: DNA hypermethylation contributes to incomplete synthesis of carbohydrate determinants in gastrointestinal cancer. Gastroenterology **135**(1), 142–151 (2008)
29. Ide, Y., Miyoshi, E., Nakagawa, T., Gu, J., Tanemura, M., Nishida, T., Ito, T., Yamamoto, H., Kozutsumi, Y., Taniguchi, N.: Aberrant expression of N-acetylglucosaminyltransferase-IVa and IVb (GnT-IVa and b) in pancreatic cancer. Biochem. Biophys. Res. Commun. **341**(2), 478–482 (2006)
30. Serpa, J., Mesquita, P., Mendes, N., Oliveira, C., Almeida, R., Santos-Silva, F., Reis, C.A., LePendu, J., David, L.: Expression of Lea in gastric cancer cell lines depends on FUT3 expression regulated by promoter methylation. Cancer Lett. **242**(2), 191–197 (2006)
31. Moriwaki, K., Narisada, M., Imai, T., Shinzaki, S., Miyoshi, E.: The effect of epigenetic regulation of fucosylation on TRAIL-induced apoptosis. Glycoconj. J. **27**(7–9), 649–659 (2010)
32. Zoldoš, V., Horvat, T., Novokmet, M., Cuenin, C., Mužinić, A., Pučić, M., Huffman, J.E., Gornik, O., Polašek, O., Campbell, H., Hayward, C., Wright, A.F., Rudan, I., Owen, K., McCarthy, M.I.,

Herceg, Z., Lauc, G.: Epigenetic silencing of HNF1A associates with changes in the composition of the human plasma N-glycome. Epigenetics **7**(2), 164–172 (2012)
33. Kaati, G., Bygren, L.O., Pembrey, M., Sjostrom, M.: Transgenerational response to nutrition, early life circumstances and longevity. Eur. J. Hum. Genet. **15**(7), 784–790 (2007)
34. Franklin, T.B., Russig, H., Weiss, I.C., Graff, J., Linder, N., Michalon, A., Vizi, S., Mansuy, I.M.: Epigenetic transmission of the impact of early stress across generations. Biol. Psychiatry **68**(5), 408–415 (2010)
35. Yauk, C., Polyzos, A., Rowan-Carroll, A., Somers, C.M., Godschalk, R.W., Van Schooten, F.J., Berndt, M.L., Pogribny, I.P., Koturbash, I., Williams, A., Douglas, G.R., Kovalchuk, O.: Germ-line mutations, DNA damage, and global hypermethylation in mice exposed to particulate air pollution in an urban/industrial location. Proc. Natl. Acad. Sci. U.S.A. **105**(2), 605–610 (2008)
36. Knežević, A., Polašek, O., Gornik, O., Rudan, I., Campbell, H., Hayward, C., Wright, A., Kolčić, I., O'Donoghue, N., Bones, J., Rudd, P.M., Lauc, G.: Variability, heritability and environmental determinants of human plasma N-glycome. J. Proteome Res. **8**, 694–701 (2009)
37. Lauc, G., Essafi, A., Huffman, J.E., Hayward, C., Knežević, A., Kattla, J.J., Polašek, O., Gornik, O., Vitart, V., Abrahams, J.L., Pučić, M., Novokmet, M., Redžić, I., Campbell, S., Wild, S.H., Borovečki, F., Wang, W., Kolčić, I., Zgaga, L., Gyllensten, U., Wilson, J.F., Wright, A.F., Hastie, N.D., Campbell, H., Rudd, P.M., Rudan, I.: Genomics meets glycomics—the first GWAS study of human N-glycome identifies HNF1alpha as a master regulator of plasma protein fucosylation. PLoS Genet. **6**(12), e1001256 (2010)
38. Huffman, J.E., Knezevic, A., Vitart, V., Kattla, J., Adamczyk, B., Novokmet, M., Igl, W., Pucic, M., Zgaga, L., Johannson, A., Redzic, I., Gornik, O., Zemunik, T., Polasek, O., Kolcic, I., Pehlic, M., Koeleman, C.A., Campbell, S., Wild, S.H., Hastie, N.D., Campbell, H., Gyllensten, U., Wuhrer, M., Wilson, J.F., Hayward, C., Rudan, I., Rudd, P.M., Wright, A.F., Lauc, G.: Polymorphisms in B3GAT1, SLC9A9 and MGAT5 are associated with variation within the human plasma N-glycome of 3533 European adults. Hum. Mol. Genet. **20**(24), 5000–5011 (2011)
39. Knežević, A., Gornik, O., Polašek, O., Pučić, M., Novokmet, M., Redžić, I., Rudd, P.M., Wright, A.F., Campbell, H., Rudan, I., Lauc, G.: Effects of aging, body mass index, plasma lipid profiles, and smoking on human plasma N-glycans. Glycobiology **20**(8), 959–969 (2010)
40. Pucic, M., Knezevic, A., Vidic, J., Adamczyk, B., Novokmet, M., Polasek, O., Gornik, O., Supraha-Goreta, S., Wormald, M.R., Redzic, I., Campbell, H., Wright, A., Hastie, N.D., Wilson, J.F., Rudan, I., Wuhrer, M., Rudd, P.M., Josic, D., Lauc, G.: High throughput isolation and glycosylation analysis of IgG-variability and heritability of the IgG glycome in three isolated human populations. Mol. Cell. Proteomics **10**(10), M111 010090 (2011)
41. Gornik, O., Pavic, T., Lauc, G.: Alternative glycosylation modulates function of IgG and other proteins—implications on evolution and disease. Biochim. Biophys. Acta. (2011). doi:10.1016/j.bbagen.2011.12.004
42. Dennis, J.W., Lau, K.S., Demetriou, M., Nabi, I.R.: Adaptive regulation at the cell surface by N-glycosylation. Traffic **10**(11), 1569–1578 (2009)
43. Dennis, J.W., Nabi, I.R., Demetriou, M.: Metabolism, cell surface organization, and disease. Cell **139**(7), 1229–1241 (2009)
44. Park, C., Zhang, J.: Genome-wide evolutionary conservation of N-glycosylation sites. Mol. Biol. Evol. **28**(8), 2351–2357 (2011)
45. Kutalik, Z., Benyamin, B., Bergmann, S., Mooser, V., Waeber, G., Montgomery, G.W., Martin, N.G., Madden, P.A., Heath, A.C., Beckmann, J.S., Vollenweider, P., Marques-Vidal, P., Whitfield, J.B.: Genome-wide association study identifies two loci strongly affecting transferrin glycosylation. Hum. Mol. Genet. **20**(18), 3710–3717 (2011)
46. Mitsumoto, Y., Oka, S., Sakuma, H., Inazawa, J., Kawasaki, T.: Cloning and chromosomal mapping of human glucuronyltransferase involved in biosynthesis of the HNK-1 carbohydrate epitope. Genomics **65**(2), 166–173 (2000)
47. Oka, S., Terayama, K., Kawashima, C., Kawasaki, T.: A novel glucuronyltransferase in nervous system presumably associated with the biosynthesis of HNK-1 carbohydrate epitope on glycoproteins. J. Biol. Chem. **267**(32), 22711–22714 (1992)
48. Roxrud, I., Raiborg, C., Gilfillan, G.D., Stromme, P., Stenmark, H.: Dual degradation mechanisms ensure disposal of NHE6 mutant protein associated with neurological disease. Exp. Cell Res. **315** (17), 3014–3027 (2009)
49. Bos, P.D., Zhang, X.H., Nadal, C., Shu, W., Gomis, R.R., Nguyen, D.X., Minn, A.J., van de Vijver, M.J., Gerald, W.L., Foekens, J.A., Massague, J.: Genes that mediate breast cancer metastasis to the brain. Nature **459**(7249), 1005–1009 (2009)
50. Brynedal, B., Wojcik, J., Esposito, F., Debailleul, V., Yaouanq, J., Martinelli-Boneschi, F., Edan, G., Comi, G., Hillert, J., Abderrahim, H.: MGAT5 alters the severity of multiple sclerosis. J. Neuroimmunol. **220**(1-2), 120–124 (2010)
51. Partridge, E.A., Le Roy, C., Di Guglielmo, G.M., Pawling, J., Cheung, P., Granovsky, M., Nabi, I.R., Wrana, J.L., Dennis, J.W.: Regulation of cytokine receptors by Golgi N-glycan processing and endocytosis. Science **306**(5693), 120–124 (2004)
52. Zhao, J., Patwa, T.H., Lubman, D.M., Simeone, D.M.: Protein biomarkers in cancer: natural glycoprotein microarray approaches. Curr. Opin. Mol. Ther. **10**(6), 602–610 (2008)
53. Lebrilla, C.B., An, H.J.: The prospects of glycan biomarkers for the diagnosis of diseases. Mol. Biosyst. **5**(1), 17–20 (2009)
54. Reis, C.A., Osorio, H., Silva, L., Gomes, C., David, L.: Alterations in glycosylation as biomarkers for cancer detection. J. Clin. Pathol. **63**(4), 322–329 (2010)
55. Lu, J.P., Knezevic, A., Wang, Y.X., Rudan, I., Campbell, H., Zou, Z.K., Lan, J., Lai, Q.X., Wu, J.J., He, Y., Song, M.S., Zhang, L., Lauc, G., Wang, W.: Screening novel biomarkers for metabolic syndrome by profiling human plasma N-glycans in Chinese Han and Croatian populations. J. Proteome Res. **10**(11), 4959–4969 (2011)
56. Ohtsubo, K., Chen, M.Z., Olefski, J.M., Marth, J.D.: Pathway to diabetes through attenuation of pancreatic beta cell glycosylation and glucose transport. Nat. Med. **17**(9), 1067–1075 (2011)
57. Kamatani, Y., Matsuda, K., Okada, Y., Kubo, M., Hosono, N., Daigo, Y., Nakamura, Y., Kamatani, N.: Genome-wide association study ofhematological and biochemical traits in a Japanese population. Nat. Genet. **42**(3), 210–215 (2010)
58. Yuan, X., Waterworth, D., Perry, J.R., Lim, N., Song, K., Chambers, J.C., Zhang, W., Vollenweider, P., Stirnadel, H., Johnson, T., Bergmann, S., Beckmann, N.D., Li, Y., Ferrucci, L., Melzer, D., Hernandez, D., Singleton, A., Scott, J., Elliott, P., Waeber, G., Cardon, L., Frayling, T.M., Kooner, J.S., Mooser, V.: Population-based genome-wide association studies reveal six loci influencing plasma levels of liver enzymes. Am. J. Hum. Genet. **83**(4), 520–528 (2008)
59. Chambers, J.C., Zhang, W., Sehmi, J., Li, X., Wass, M.N., Van der Harst, P., Holm, H., Sanna, S., Kavousi, M., Baumeister, S.E., Coin, L.J., Deng, G., Gieger, C., Heard-Costa, N.L., Hottenga, J.J., Kuhnel, B., Kumar, V., Lagou, V., Liang, L., Luan, J., Vidal, P.M., Mateo Leach, I., O'Reilly, P.F., Peden, J.F., Rahmioglu, N., Soininen, P., Speliotes, E.K., Yuan, X., Thorleifsson, G., Alizadeh, B.Z., Atwood, L.D., Borecki, I.B., Brown, M.J., Charoen, P., Cucca, F., Das, D., de Geus, E.J., Dixon, A.L., Doring, A., Ehret, G., Eyjolfsson, G.I., Farrall, M., Forouhi, N.G., Friedrich, N., Goessling, W., Gudbjartsson, D.F., Harris, T.B., Hartikainen, A.L., Heath, S., Hirschfield, G.M., Hofman, A., Homuth, G., Hypponen, E., Janssen, H.L., Johnson, T., Kangas, A.J., Kema, I.P., Kuhn, J.P., Lai, S., Lathrop, M., Lerch, M.M., Li, Y., Liang,

T.J., Lin, J.P., Loos, R.J., Martin, N.G., Moffatt, M.F., Montgomery, G.W., Munroe, P.B., Musunuru, K., Nakamura, Y., O'Donnell, C.J., Olafsson, I., Penninx, B.W., Pouta, A., Prins, B.P., Prokopenko, I., Puls, R., Ruokonen, A., Savolainen, M.J., Schlessinger, D., Schouten, J.N., Seedorf, U., Sen-Chowdhry, S., Siminovitch, K.A., Smit, J.H., Spector, T.D., Tan, W., Teslovich, T.M., Tukiainen, T., Uitterlinden, A.G., Van der Klauw, M.M., Vasan, R.S., Wallace, C., Wallaschofski, H., Wichmann, H.E., Willemsen, G., Wurtz, P., Xu, C., Yerges-Armstrong, L.M., Abecasis, G.R., Ahmadi, K.R., Boomsma, D.I., Caulfield, M., Cookson, W.O., van Duijn, C.M., Froguel, P., Matsuda, K., McCarthy, M.I., Meisinger, C., Mooser, V., Pietilainen, K.H., Schumann, G., Snieder, H., Sternberg, M.J., Stolk, R.P., Thomas, H.C., Thorsteinsdottir, U., Uda, M., Waeber, G., Wareham, N.J., Waterworth, D.M., Watkins, H., Whitfield, J.B., Witteman, J.C., Wolffenbuttel, B.H., Fox, C.S., Ala-Korpela, M., Stefansson, K., Vollenweider, P., Volzke, H., Schadt, E.E., Scott, J., Jarvelin, M.R., Elliott, P., Kooner, J.S.: Genome-wide association study identifies loci influencing concentrations of liver enzymes in plasma. Nat. Genet. **43**(11), 1131–1138 (2011)

60. Chung, C.M., Wang, R.Y., Chen, J.W., Fann, C.S., Leu, H.B., Ho, H.Y., Ting, C.T., Lin, T.H., Sheu, S.H., Tsai, W.C., Chen, J.H., Jong, Y.S., Lin, S.J., Chen, Y.T., Pan, W.H.: A genome-wide association study identifies new loci for ACE activity: potential implications for response to ACE inhibitor. Pharmacogenomics J. **10**(6), 537–544 (2010)
61. Teupser, D., Baber, R., Ceglarek, U., Scholz, M., Illig, T., Gieger, C., Holdt, L.M., Leichtle, A., Greiser, K.H., Huster, D., Linsel-Nitschke, P., Schafer, A., Braund, P.S., Tiret, L., Stark, K., Raaz-Schrauder, D., Fiedler, G.M., Wilfert, W., Beutner, F., Gielen, S., Grosshennig, A., Konig, I.R., Lichtner, P., Heid, I.M., Kluttig, A., El Mokhtari, N.E., Rubin, D., Ekici, A.B., Reis, A., Garlichs, C.D., Hall, A.S., Matthes, G., Wittekind, C., Hengstenberg, C., Cambien, F., Schreiber, S., Werdan, K., Meitinger, T., Loeffler, M., Samani, N.J., Erdmann, J., Wichmann, H.E., Schunkert, H., Thiery, J.: Genetic regulation of serum phytosterol levels and risk of coronary artery disease. Circ. Cardiovasc. Genet. **3**(4), 331–339 (2010)
62. Qi, L., Cornelis, M.C., Kraft, P., Jensen, M., van Dam, R.M., Sun, Q., Girman, C.J., Laurie, C.C., Mirel, D.B., Hunter, D.J., Rimm, E., Hu, F.B.: Genetic variants in ABO blood group region, plasma soluble E-selectin levels and risk of type 2 diabetes. Hum. Mol. Genet. **19**(9), 1856–1862 (2010)
63. Paterson, A.D., Lopes-Virella, M.F., Waggott, D., Boright, A.P., Hosseini, S.M., Carter, R.E., Shen, E., Mirea, L., Bharaj, B., Sun, L., Bull, S.B.: Genome-wide association identifies the ABO blood group as a major locus associated with serum levels of soluble E-selectin. Arterioscler. Thromb. Vasc. Biol. **29**(11), 1958–1967 (2009)
64. Barbalic, M., Dupuis, J., Dehghan, A., Bis, J.C., Hoogeveen, R.C., Schnabel, R.B., Nambi, V., Bretler, M., Smith, N.L., Peters, A., Lu, C., Tracy, R.P., Aleksic, N., Heeriga, J., Keaney Jr., J.F., Rice, K., Lip, G.Y., Vasan, R.S., Glazer, N.L., Larson, M.G., Uitterlinden, A.G., Yamamoto, J., Durda, P., Haritunians, T., Psaty, B.M., Boerwinkle, E., Hofman, A., Koenig, W., Jenny, N.S., Witteman, J.C., Ballantyne, C., Benjamin, E.J.: Large-scale genomic studies reveal central role of ABO in sP-selectin and sICAM-1 levels. Hum. Mol. Genet. **19**(9), 1863–1872 (2010)
65. Pare, G., Chasman, D.I., Kellogg, M., Zee, R.Y., Rifai, N., Badola, S., Miletich, J.P., Ridker, P.M.: Novel association of ABO histo-blood group antigen with soluble ICAM-1: results of a genome-wide association study of 6,578 women. PLoS Genet. **4**(7), e1000118 (2008)
66. Melzer, D., Perry, J.R., Hernandez, D., Corsi, A.M., Stevens, K., Rafferty, I., Lauretani, F., Murray, A., Gibbs, J.R., Paolisso, G., Rafiq, S., Simon-Sanchez, J., Lango, H., Scholz, S., Weedon, M.N., Arepalli, S., Rice, N., Washecka, N., Hurst, A., Britton, A., Henley, W., van de Leemput, J., Li, R., Newman, A.B., Tranah, G., Harris, T., Panicker, V., Dayan, C., Bennett, A., McCarthy, M.I., Ruokonen, A., Jarvelin, M.R., Guralnik, J., Bandinelli, S., Frayling, T.M., Singleton, A., Ferrucci, L.: A genome-wide association study identifies protein quantitative trait loci (pQTLs). PLoS Genet. **4**(5), e1000072 (2008)
67. Tregouet, D.A., Heath, S., Saut, N., Biron-Andreani, C., Schved, J.F., Pernod, G., Galan, P., Drouet, L., Zelenika, D., Juhan-Vague, I., Alessi, M.C., Tiret, L., Lathrop, M., Emmerich, J., Morange, P.E.: Common susceptibility alleles are unlikely to contribute as strongly as the FV and ABO loci to VTE risk: results from a GWAS approach. Blood **113**(21), 5298–5303 (2009)
68. Johansson, A., Marroni, F., Hayward, C., Franklin, C.S., Kirichenko, A.V., Jonasson, I., Hicks, A.A., Vitart, V., Isaacs, A., Axenovich, T., Campbell, S., Floyd, J., Hastie, N., Knott, S., Lauc, G., Pichler, I., Rotim, K., Wild, S.H., Zorkoltseva, I.V., Wilson, J.F., Rudan, I., Campbell, H., Pattaro, C., Pramstaller, P., Oostra, B.A., Wright, A.F., van Duijn, C.M., Aulchenko, Y.S., Gyllensten, U.: Linkage and genome-wide association analysis of obesity-related phenotypes: association of weight with the MGAT1 gene. Obesity (Silver Spring) **18**(4), 803–808 (2010)
69. Daly, A.K., Donaldson, P.T., Bhatnagar, P., Shen, Y., Pe'er, I., Floratos, A., Daly, M.J., Goldstein, D.B., John, S., Nelson, M.R., Graham, J., Park, B.K., Dillon, J.F., Bernal, W., Cordell, H.J., Pirmohamed, M., Aithal, G.P., Day, C.P.: HLA-B*5701 genotype is a major determinant of drug-induced liver injury due to flucloxacillin. Nat. Genet. **41**(7), 816–819 (2009)
70. Franke, A., McGovern, D.P., Barrett, J.C., Wang, K., Radford-Smith, G.L., Ahmad, T., Lees, C.W., Balschun, T., Lee, J., Roberts, R., Anderson, C.A., Bis, J.C., Bumpstead, S., Ellinghaus, D., Festen, E.M., Georges, M., Green, T., Haritunians, T., Jostins, L., Latiano, A., Mathew, C.G., Montgomery, G.W., Prescott, N.J., Raychaudhuri, S., Rotter, J.I., Schumm, P., Sharma, Y., Simms, L.A., Taylor, K.D., Whiteman, D., Wijmenga, C., Baldassano, R.N., Barclay, M., Bayless, T.M., Brand, S., Buning, C., Cohen, A., Colombel, J.F., Cottone, M., Stronati, L., Denson, T., De Vos, M., D'Inca, R., Dubinsky, M., Edwards, C., Florin, T., Franchimont, D., Gearry, R., Glas, J., Van Gossum, A., Guthery, S.L., Halfvarson, J., Verspaget, H.W., Hugot, J.P., Karban, A., Laukens, D., Lawrance, I., Lemann, M., Levine, A., Libioulle, C., Louis, E., Mowat, C., Newman, W., Panes, J., Phillips, A., Proctor, D.D., Regueiro, M., Russell, R., Rutgeerts, P., Sanderson, J., Sans, M., Seibold, F., Steinhart, A.H., Stokkers, P.C., Torkvist, L., Kullak-Ublick, G., Wilson, D., Walters, T., Targan, S.R., Brant, S.R., Rioux, J.D., D'Amato, M., Weersma, R.K., Kugathasan, S., Griffiths, A.M., Mansfield, J.C., Vermeire, S., Duerr, R.H., Silverberg, M.S., Satsangi, J., Schreiber, S., Cho, J.H., Annese, V., Hakonarson, H., Daly, M.J., Parkes, M.: Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. Nat. Genet. **42**(12), 1118–1125 (2010)
71. McGovern, D.P., Jones, M.R., Taylor, K.D., Marciante, K., Yan, X., Dubinsky, M., Ippoliti, A., Vasiliauskas, E., Berel, D., Derkowski, C., Dutridge, D., Fleshner, P., Shih, D.Q., Melmed, G., Mengesha, E., King, L., Pressman, S., Haritunians, T., Guo, X., Targan, S.R., Rotter, J.I.: Fucosyltransferase 2 (FUT2) non-secretor status is associated with Crohn's disease. Hum. Mol. Genet. **19**(17), 3468–3476 (2010)
72. Tanaka, T., Scheet, P., Giusti, B., Bandinelli, S., Piras, M.G., Usala, G., Lai, S., Mulas, A., Corsi, A.M., Vestrini, A., Sofi, F., Gori, A.M., Abbate, R., Guralnik, J., Singleton, A., Abecasis, G.R., Schlessinger, D., Uda, M., Ferrucci, L.: Genome-wide association study of vitamin B6, vitamin B12, folate, and homocysteine blood concentrations. Am. J. Hum. Genet. **84**(4), 477–482 (2009)
73. Hazra, A., Kraft, P., Selhub, J., Giovannucci, E.L., Thomas, G., Hoover, R.N., Chanock, S.J., Hunter, D.J.: Common variants of

FUT2 are associated with plasma vitamin B12 levels. Nat. Genet. **40**(10), 1160–1162 (2008)

74. Willer, C.J., Sanna, S., Jackson, A.U., Scuteri, A., Bonnycastle, L.L., Clarke, R., Heath, S.C., Timpson, N.J., Najjar, S.S., Stringham, H.M., Strait, J., Duren, W.L., Maschio, A., Busonero, F., Mulas, A., Albai, G., Swift, A.J., Morken, M.A., Narisu, N., Bennett, D., Parish, S., Shen, H., Galan, P., Meneton, P., Hercberg, S., Zelenika, D., Chen, W.M., Li, Y., Scott, L.J., Scheet, P.A., Sundvall, J., Watanabe, R.M., Nagaraja, R., Ebrahim, S., Lawlor, D.A., Ben-Shlomo, Y., Davey-Smith, G., Shuldiner, A.R., Collins, R., Bergman, R.N., Uda, M., Tuomilehto, J., Cao, A., Collins, F.S., Lakatta, E., Lathrop, G.M., Boehnke, M., Schlessinger, D., Mohlke, K.L., Abecasis, G.R.: Newly identified loci that influence lipid concentrations and risk of coronary artery disease. Nat. Genet. **40**(2), 161–169 (2008)
75. Kathiresan, S., Melander, O., Guiducci, C., Surti, A., Burtt, N.P., Rieder, M.J., Cooper, G.M., Roos, C., Voight, B.F., Havulinna, A.S., Wahlstrand, B., Hedner, T., Corella, D., Tai, E.S., Ordovas, J.M., Berglund, G., Vartiainen, E., Jousilahti, P., Hedblad, B., Taskinen, M.R., Newton-Cheh, C., Salomaa, V., Peltonen, L., Groop, L., Altshuler, D.M., Orho-Melander, M.: Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. Nat. Genet. **40**(2), 189–197 (2008)
76. Waterworth, D.M., Ricketts, S.L., Song, K., Chen, L., Zhao, J.H., Ripatti, S., Aulchenko, Y.S., Zhang, W., Yuan, X., Lim, N., Luan, J., Ashford, S., Wheeler, E., Young, E.H., Hadley, D., Thompson, J.R., Braund, P.S., Johnson, T., Struchalin, M., Surakka, I., Luben, R., Khaw, K.T., Rodwell, S.A., Loos, R.J., Boekholdt, S.M., Inouye, M., Deloukas, P., Elliott, P., Schlessinger, D., Sanna, S., Scuteri, A., Jackson, A., Mohlke, K.L., Tuomilehto, J., Roberts, R., Stewart, A., Kesaniemi, Y.A., Mahley, R.W., Grundy, S.M., McArdle, W., Cardon, L., Waeber, G., Vollenweider, P., Chambers, J.C., Boehnke, M., Abecasis, G.R., Salomaa, V., Jarvelin, M.R., Ruokonen, A., Barroso, I., Epstein, S.E., Hakonarson, H.H., Rader, D.J., Reilly, M.P., Witteman, J.C., Hall, A.S., Samani, N.J., Strachan, D.P., Barter, P., van Duijn, C.M., Kooner, J.S., Peltonen, L., Wareham, N.J., McPherson, R., Mooser, V., Sandhu, M.S.: Genetic variants influencing circulating lipid levels and risk of coronary artery disease. Arterioscler. Thromb. Vasc. Biol. **30** (11), 2264–2276 (2010)
77. Reveille, J.D., Sims, A.M., Danoy, P., Evans, D.M., Leo, P., Pointon, J.J., Jin, R., Zhou, X., Bradbury, L.A., Appleton, L.H., Davis, J.C., Diekman, L., Doan, T., Dowling, A., Duan, R., Duncan, E.L., Farrar, C., Hadler, J., Harvey, D., Karaderi, T., Mogg, R., Pomeroy, E., Pryce, K., Taylor, J., Savage, L., Deloukas, P., Kumanduri, V., Peltonen, L., Ring, S.M., Whittaker, P., Glazov, E., Thomas, G.P., Maksymowych, W.P., Inman, R.D., Ward, M.M., Stone, M.A., Weisman, M.H., Wordsworth, B.P., Brown, M.A.: Genome-wide association study of ankylosing spondylitis identifies non-MHC susceptibility loci. Nat. Genet. **42**(2), 123–127 (2010)
78. Schaefer, A.S., Richter, G.M., Nothnagel, M., Manke, T., Dommisch, H., Jacobs, G., Arlt, A., Rosenstiel, P., Noack, B., Groessner-Schreiber, B., Jepsen, S., Loos, B.G., Schreiber, S.: A genome-wide association study identifies GLT6D1 as a susceptibility locus for periodontitis. Hum. Mol. Genet. **19**(3), 553–562 (2010)
79. Gudbjartsson, D.F., Walters, G.B., Thorleifsson, G., Stefansson, H., Halldorsson, B.V., Zusmanovich, P., Sulem, P., Thorlacius, S., Gylfason, A., Steinberg, S., Helgadottir, A., Ingason, A., Steinthorsdottir, V., Olafsdottir, E.J., Olafsdottir, G.H., Jonsson, T., Borch-Johnsen, K., Hansen, T., Andersen, G., Jorgensen, T., Pedersen, O., Aben, K.K., Witjes, J.A., Swinkels, D.W., den Heijer, M., Franke, B., Verbeek, A.L., Becker, D.M., Yanek, L.R., Becker, L.C., Tryggvadottir, L., Rafnar, T., Gulcher, J., Kiemeney, L.A., Kong, A., Thorsteinsdottir, U., Stefansson, K.: Many sequence variants affecting diversity of adult human height. Nat. Genet. **40**(5), 609–615 (2008)