HIGH THROUGHPUT MULTIPLEX SNP-ANALYSIS IN CHRONIC OBSTRUCTIVE PULMONARY DISEASE AND LUNG CANCER

Zsuzsanna Elek[1], Zsuzsanna Kovács[2], Gergely Keszler[1], Miklós Szabó[3], Eszter Csanky[3], Jane Luo[4], András Guttman[2,4,5], Zsolt Rónai[1,*]

[1] Department of Medical Chemistry, Molecular Biology and Pathobiochemistry, Semmelweis University, Budapest, Hungary
[2] Horváth Csaba Memorial Laboratory of Bioseparation Sciences, Research Center for Molecular Medicine, Doctoral School of Molecular Medicine, Faculty of Medicine, University of Debrecen, 98 Nagyerdei krt., Debrecen, 4032, Hungary
[3] Semmelweis Hospital, Miskolc, Hungary
[4] SCIEX Separations, Brea, CA 92821, USA
[5] Translational Glycomics Group, Research Institute of Biomolecular and Chemical Engineering, University of Pannonia, Veszprém, Hungary

*corresponding author: Zsolt Rónai
Department of Medical Chemistry, Molecular Biology and Pathobiochemistry, Semmelweis University, H-1094 Budapest, Tűzoltó utca 37–47., Hungary
e-mail: ronai.zsolt@med.semmelweis-univ.hu

Keywords: SNP, single base primer extension, capillary gel electrophoresis, TaqMan OpenArray, lung adenocarcinoma, COPD, genetic association

Running Title: Multiplex SNP-Analysis in Pulmonary Diseases

ABSTRACT

Background

A number of human inflammatory diseases and tumors have been shown to cause alterations in the glycosylation pattern of plasma proteins in a specific manner. These highly variable and versatile post-translational modifications fine-tune protein functions by influencing sorting, folding, enzyme activity and subcellular localization. However, relatively little is known about regulatory factors of this procedure and about the accurate causative connection between glycosylation and disease.

Objective

The aim of the present study was to investigate whether certain single nucleotide polymorphisms (SNPs) in genes encoding glycosyltransferases and glycosidases could be associated with elevated risk for chronic obstructive pulmonary disease (COPD) and lung adenocarcinoma.

Methods

A total of 32 SNPs localized in genes related to N-glycosylation were selected for the association analysis. Polymorphisms with putative biological functions (missense or regulatory variants) were recruited. SNPs were genotyped by a TaqMan OpenArray platform. A single base extension based method in combination with capillary gel electrophoresis was used for verification.

Results

The TaqMan OpenArray approach provided accurate and reliable genotype data (global call rate: 94.9%, accuracy: 99.6%). No significant discrepancy was detected between the obtained and expected genotype frequency values (Hardy–Weinberg equilibrium) in the healthy control sample group in case of any SNP confirming reliable sampling and genotyping. Allele frequencies of the rs3944508 polymorphism localized in the 3’ UTR of the MGAT5 gene significantly differed in the sample groups compared.

Conclusion

Our results suggest that the rs34944508 SNP might modulate the risk for lung cancer by influencing the expression of MGAT5. This enzyme catalyzes the addition of N-acetylglucosamine (GlcNAc) in beta 1-6 linkage to the alpha-linked mannose of biantennary N-linked oligosaccharides, thus, increases branching that characteristic for invasive malignancies.

INTRODUCTION

Protein glycosylation has long been known as one of the most common, highly variable and versatile post-translational modifications that fine-tunes protein functions by modulating sorting, folding, enzyme activity and subcellular localization [1, 2]. The astounding diversity of protein-attached glycans is the result of the concerted action of glycosyltransferases and glycosidases, which function is template-independent. To date, however, relatively little is known about the regulatory factors, which coordinate the tissue-specific establishment, maintenance and dynamic modulation of protein glycosylation patterns across the body [3, 4]. The advent of high-throughput analytical methods such as reversed phase liquid chromatography and capillary electrophoresis, especially when connected to mass spectroscopy have shed light on marked individual variability of the glycoproteome [4-6]. It also turned out that cells respond to various environmental effects by dynamically altering their global glycosylation patterns, and recent advances in the burgeoning field of glycoproteomics have contributed a great deal to our better understanding of disease-associated alterations of the glycome [7, 8].

A number of human inflammatory diseases and tumors seems to influence the glycosylation patterns of plasma proteins in a specific manner, making them potential candidates in the quest to find reliable and easily accessible biomarkers [9, 10]. N-glycan profiling of plasma proteins unveiled characteristic alterations both in systemic [11, 12] and organ-specific inflammation [13], and the diagnosis of pancreatic [14, 15], breast [16, 17] and ovarian [18, 19] cancer might be corroborated by early N-glycosylation analysis. Moreover, serum N-glycans might help to discriminate between benign and malignant tumors [20], predict the prognosis in terms of recurrence and metastasis [21] and the propagation of chronic viral hepatitis to cirrhosis and eventually hepatocellular carcinoma [22]. Another relevant example for the inflammation-to-malignancy transition is chronic obstructive pulmonary disease (COPD), a chronic airway inflammation, which is a common risk factor for lung cancer with overlapping genetic background [23]. As of today, several serum glycan biomarkers predictive for lung cancer have been described [24-26], and we recently reported that differences in fucosylation levels and positional isomer types of plasma haptoglobin can help to distinguish between COPD and pulmonary cancer [27].

Cancer-specific disturbances in N-glycan profiles might emerge due to genomic sequence variants occurring either in the coding or regulatory regions of key glycosylation enzymes [28]. Length or point polymorphisms in promoters or in 3' untranslated regions can result in subtle changes in their expression levels by modulating transcription factor or miRNA binding affinities [29, 30], respectively, while missense mutations often lead to altered activities or substrate specificities. Theoretically, further disturbances in the glycoproteome might arise due to variations in consensus glycosylation sequences in target proteins [31] or altered synthesis or cellular uptake of precursor carbohydrates. However, it is still unclear

whether altered glycosylation patterns play an indispensable role in the acquisition of cancer-specific hallmarks, or they occur as mere consequences of the malignant transformation [32].

The aim of the present study was to clarify whether certain single nucleotide polymorphisms (SNPs) in genes encoding glycosyltransferases and glycosidases could be associated with elevated risk for chronic obstructive pulmonary disease (COPD) and lung adenocarcinoma. To this end, cohorts of COPD, lung cancer and co-morbid patients as well as healthy controls were genotyped using two different methodologies, a high-throughput, TaqMan probe based OpenArray platform approach [33] and primer extension combined with capillary electrophoresis [34].

MATERIALS AND METHODS

Participants, DNA-sampling and -purification

Non-related Hungarian (Caucasian) male and female subjects (age: 40–85 years) participated in our study. People of different ethnic origin were excluded from the analyses to avoid population stratification. Participants signed written informed consent and the study protocol was approved by the Scientific and Research Ethics Committee of the Medical Research Council (ETT TUKEB ad.328/KO/2005, ad.323-86/2005-1018EKU).

Participants suffering from COPD or bronchial cancer were selected from patients treated in "Borsod-Abaúj-Zemplén" County Central Hospital (Miskolc, Hungary). Patients with other respiratory disease, such as asthma, sarcoidosis, tuberculosis, lung fibrosis, cystic fibrosis, as well as any tumor or oncological treatment in the past 5 years were excluded from the study. COPD was diagnosed according to the protocol of GOLD 2016 (Global initiative of Obstructive Lung Disease) based on the result of spirometry examination. Diagnosis of COPD was stated if the FEV1 / FVC ratio (Tiffeneau-Pinelli index – proportion of forced expiratory volume in the first second to the forced vital capacity) was less than 0.7. Current spirometry data were taken into consideration in case of patients with stable condition, whereas earlier data were used for subjects with acute exacerbating COPD. Subjects of the lung cancer group were selected from patients with recent diagnosis confirmed by histology or cytology, who has not received any antitumor therapy yet. Members of the control group were volunteers with similar smoking habit, however, without any pulmonary disease in their medical history. Based on these aspects, subjects were categorized into the following four groups: (1) healthy control (N = 294), (2) COPD (N = 308), (3) lung cancer (N = 312) and (4) comorbid COPD and lung cancer (N = 179). The four patient groups were matched according to age (51–78 years old, average: 64) and sex (male: 52%, female: 48%).

The study was approved by the Scientific and Research Ethics Committee of the Medical Research Council, ETT TUKEB ad.328/KO/2005, ad.323-86/2005-1018EKU, Hungary. No animals were used in this study, reported experiments on humans were followed in accordance with the ethical standards of the committee responsible for human

experimentation (Scientific and Research Ethics Committee of the Medical Research Council – "ETT TUKEB"), and with the Helsinki Declaration of 1975, as revised in 2008 (http://www.wma.net/en/20activities/10ethics/10helsinki/). Written consent was obtained from the individuals.

DNA-samples were collected by rubbing epithelial cells from the inner surface of the cheek or gum using cotton swabs. DNA isolation was initiated by cell lysis in a 0.2 mg/mL Proteinase K buffer solution (56 °C, 3 hours or overnight). Proteins were precipitated by the addition of saturated NaCl. After centrifugation (13,000 rpm, 20 min, 4 °C) DNA was precipitated from the supernatant by isopropanol, then pellet was washed by 70% ethanol. Finally, DNA was resolved in 0.5× TE buffer solution (5 mM Tris-HCl, pH 8.0; 0.5 mM EDTA). Concentration of the DNA-samples was measured by NanoDrop1000 spectrophotometer (Thermo Scientific, Waltham, MA, USA).

Multiplex SNP Analysis by Single Base Primer Extension and Denaturing CE

The first step of the genotype analyses was the amplification of the DNA-sequences flanking the SNPs of interest. PCR-primers were designed by NCBI Primer-Blast (https://www.ncbi.nlm.nih.gov/tools/primer-blast/). Data about each PCR primer are summarized in *Table 1*. The Qiagen HotStarTaq DNA-polymerase kit including the DNA-polymerase, a reaction buffer and a "Q-solution" was employed for amplification. Reactions were performed in a total volume of 12 µL. The flanking regions of the SNPs were amplified in multiplex PCR. Reaction mixtures contained 0.2 mM of each deoxyribonucleoside-triphosphate, 0.1 µM of the appropriate primers, 0.3 U of HotStarTaq DNA-polymerase, 1× buffer, 1× Q-solution and approximately 5 ng of genomic DNA. The first step of the thermocycle was the initial denaturation at 95 °C for 15 min to activate the HotStarTaq DNA-polymerase. It was followed by 35 cycles of denaturation (94 °C – 1 min), annealing (67 °C – 30 sec) and extension (72 °C – 1 min). PCR was completed by a final extension at 72 °C for 10 min and the samples were then kept at 8 °C until downstream analyses. Aliquots of the PCR products were analyzed by sub-marine agarose gel-electrophoresis to verify the optimal formation of each product in the multiplex PCR. Products were then treated by Exonuclease I and Shrimp Alkaline Phosphatase (New England Biolabs, Ipswich, MA) to eliminate excess primers and dNTPs according to manufacturer's protocols.

The SNPStart solution (Beckman Coulter, Brea, CA) contained the components (fluorescently labeled acyclo-terminators, DNA-polymerase and buffer) of the SBE reaction in a pre-mixed form. Reactions were carried out in 10 µL final volume, which contained 4 µL master mix, 100 fmol product of the 1st PCR and 15 pmol SBE primer. PCR consisted of 30 cycles of 90 °C denaturation for 10 sec, 55 °C annealing for 20 sec and 72 °C extension for 1 minute. Dye terminators not incorporated into the products were eliminated by Shrimp Alkaline Phosphatase (New England Biolabs) digestion according to the manufacturer's protocols. 0.5 µL of SBE reaction product was then added to 39 µL of Sample Loading Solution and 0.5 µL

of Size Standard 80 (Beckman Coulter). Samples were then covered with one drop of mineral oil. Multicapillary gel electrophoresis analysis was carried out in a GenomeLab™ GeXP Genetic Analysis System (SCIEX, Brea, CA) using the following parameters: capillary temperature: 50 °C, denaturation: 90 °C for 60 sec, sample injection 2.0 kV for 30 sec, separation: 6.0 kV for 1 min. SNP ver2 dye mobility correction was employed for data analysis.

SNP Analysis by TaqMan® OpenArray® system

The TaqMan® OpenArray® system was used for the genotype determination of a set of SNPs listed in *Table 2*. Genomic DNA samples were diluted to adjust their concentration in the range of 20–50 ng/μL. 2.5 μl of the samples were mixed with 2.5 μL TaqMan™ OpenArray™ Genotyping PCR Master Mix containing the AmpliTaq Gold® DNA-polymerase and the deoxyribonucleoside-triphosphates (dATP, dCTP, dGTP and dTTP) buffer (ThermoFisher, Waltham MA) in a 384-well-plate. This plate is used by the OpenArray™ AccuFill™ System (ThermoFisher, Waltham MA), which loads the reaction mixture on the TaqMan™ OpenArray™ Genotyping Plates (ThermoFisher). These plates were divided into 48 subarrays, each consisting of 64 through-holes allowing to perform 3072 genotype analyses (32 SNPs of 96 samples in our setup) on one single plate. Primers and sequence specific TaqMan probes were imposed on the edge of the through-holes and dissolved in the reaction mixtures after loading with DNA samples. PCRs were carried out in 33 nl reaction volume, the thermocycler was operated according to the manufacturer's OpenArray™ Genotyping protocol. Amplification, data collection during PCR as well as allelic discrimination protocol clustering the samples into 3 groups according to their genotypes were carried out by the QuantStudio™ 12K Flex Real-Time PCR platform. Genotype analysis was performed by the TaqMan Genotyper Software and by the cloud service (ThermoFisher) called "Connect". The data supporting the findings of the article are available in the ThermoFisher Connect at https://www.thermofisher.com/hu/en/home/digital-science.html, reference number 2018_CT.

Statistical Analysis

Reliability of the applied genotyping protocols was verified by testing the Hardy–Weinberg equilibrium in case of each locus. SPSS v17.0 was applied to perform χ^2-test in case–control setup to assess genotype–phenotype association. The Bonferroni correction and the False Discovery Rate (Benjamini–Hochberg procedure) were used to rule out false positive findings caused by multiplex testing.

RESULTS AND DISCUSSION

Two substantially different techniques (*Table 3*) were used for the genotype analysis of SNPs in genes (*Table 2*) coding for proteins potentially involved in the synthesis of N-glycans in

Golgi. The TaqMan® OpenArray® method used in this work was based on the simultaneous application of two probes that were labeled with two different fluorescent dyes (VIC and FAM) at their 5' end, and could anneal to the two alleles of the given locus, respectively. This allowed genotype determination of one single SNP in each reaction (multiplex analyses cannot be performed). The procedure on the other hand consisted of one single step, which in combination with the miniaturized, low density "Open Array" format offered extraordinary high throughput. The great majority of the genotyping procedures were carried out by this technique.

The other approach – also known as "minisequencing" – employed the chain terminating acyclo (9-(2-hydroxyethoxy)methyl-) nucleotides labeled with different fluorophores. Genotype determination was based on the application of a primer annealing to the nucleotide nearby the SNP site with its 3' end in the single base extension (SBE) reaction. This primer was elongated by one single acyclo nucleotide, thus the color of the product allowed unambiguous determination of the alleles present in the polymorphic locus. Detection of the generated products was carried out by capillary gel electrophoresis (CGE), offering fragment size determination as well. This allowed the completion of multiplex assays, as the application of SBE primers with different lengths resulted in reaction products that could be separated by CGE. The whole analysis consisted of three major steps: 1) first the flanking region of the investigated SNPs was amplified. This reaction can also be carried out in a multiplex format, or – in case of loci in close proximity with each other – several SNPs can be included in one PCR amplicon; 2) The product of this PCR served as a template for the SBE reaction, which 3) was followed by CGE analysis. This approach offered excellent reliability, however, – despite of the potential of carrying out multiplex measurements – the three consecutive steps were time-consuming and labor-intensive, reducing the throughput of the workflow. Thus, this technique was used for validation of the real-time PCR based approach for a subset of samples and polymorphisms.

Figure 1 shows the results of the representative genotype analyses obtained by the two methods. As one can observe, samples with different genotypes form clusters in case in the TaqMan assay (*Figure 1A*). High signal of only one of the two dyes can be observed in case of homozygotes (blue and red dots). Heterozygotes are localized in the middle of the graph (green dots). No amplification occured in the non-template control (NTC) samples, consequently they can be found close to the origin of the graph. The unambiguous representation of the three genotypes can be seen in *Figure 1B* in a single SNP analysis by primer extension in combination with high throughput, multiplex CGE. Please note, that this CGE-based method possesses outstanding specificity, as absolutely no aspecific signal can be observed in case of homozygote samples.

A total of 32 SNPs (*Table 2*) localized in the MGAT5, ST3GAL1, ST6GAL1, FUT2 and FUT3 N-glycosylation related genes were selected for the case–control analysis. Polymorphisms with putative biological functions were recruited, i.e., missense variants and loci in the 5' UTR, 1st intron or 3' UTR were involved in our study. In case of the loci localized in the 5' UTR or in the 1st intron an alteration in the efficiency in transcription factor binding was suggested by *in silico* prediction, whereas 3' UTR SNPs may often modulate the mRNA–microRNA interaction, which can lead to the alteration of translation activity. A subset of 96 samples were analyzed for the rs2922471 C/T, rs4736675 A/G and rs1042757 C/G SNPs (*Table 1*) with both methods resulting in as high as 99.6% concordant results. The one single discrepancy was also resolved after verification of the results of the two assays by reconsidering the automatically called genotype by the TaqMan® OpenArray® system. Two polymorphisms (rs2230908, rs1468906) were excluded from the downstream genotype–phenotype association study because of ambiguous genotyping data. Samples formed five distinct clusters instead of the 3 genotype categories in case of the rs2230908, which might be caused by either a copy number variation or another adjacent SNP in the region of the investigated locus. In case of the rs1468906 polymorphism the efficiency of the amplification was very poor in approximately half of the samples. It was apparently caused by a similar issue, i.e., an unknown polymorphism might have inhibited adequate binding of the primers or the probes, resulting in the lack of the PCR-product.
The Hardy–Weinberg equilibrium was used to evaluate the reliability of the genotype analysis workflow in case of each SNP. The χ^2-test did not show any significant discrepancy between the obtained and expected genotype frequency values in the healthy control sample group confirming reliable sampling and genotyping.

Results of the case–control study are summarized in *Table 4*. Analysis was carried out in an allele-wise manner, i.e., the distribution of the allele-frequency values was compared in the four patient groups (controls, patients with COPD, lung cancer and comorbid cases). Application of the allele-wise approach helped to avoid false positive findings caused by relatively rare genotype categories or by heterozygotes. Significant difference (nominal $p < 0.05$) was detected in case of the rs34944508 SNP in the MGAT5 gene and the rs35166820 polymorphism in ST6GAL1. The latter locus did not reach the level of statistical significance after carrying out corrections for multiple testing. It is notable on the other hand that the difference in the distribution of the allele frequencies among the four groups remained statistically significant in case of the rs34944508 locus using either the FDR (Benjamini–Hochberg false discovery rate) or the more stringent Bonferroni-correction. Results are visualized in *Figure 2* showing an outstanding, more than 2× difference in the minor allele frequencies of the healthy controls (8.7%) and the patients with lung cancer (3.3%), whereas the two other subject groups (COPD: 4.5% and comorbid: 4.1%) possessed intermediate values.

The 3’ untranslated regions (3’ UTRs) of genes often contain cognate binding sites for microRNA. miRNA binding may cause downregulation of translation and/or mRNA degradation. Therefore, we raised the question whether the rs34944508 in the 3’ UTR of the N-acetylglucosaminyltransferase V (MGAT5) gene might affected the miRNA binding sites. Searching in the PolymiRTS database, it turned out that the minor T allele of the SNP creates and the common C variant abrogates a binding site for the hsa-miR-3200-3p miRNA. The expression of this miRNA has been shown to be controlled by NF-κB in HeLa cells [35], and its plasma levels were found to be reduced in Alzheimer disease, making it a potential miRNA-biomarker of neurodegeneration [36]. The latter result has been recapitulated in intact red blood cells too [37].

As the C allele was found to be enriched in the lung cancer population, it seemed tempting to assume that elevated levels of MGAT5 might be involved in the pathogenesis of pulmonary cancer. Our hypothesis is underpinned by the following results. First, the miR-3200-3p has been shown to be expressed in lung tissue according to miRGator 3.0, so its potential regulatory role in lung cancer is apparently biologically relevant. Second, it was earlier reported [38] that siRNA-mediated knockdown of MGAT5 inhibited the growth of lung cancer cells both *in vitro* and *in vivo*. MGAT5 is a typical cancer-associated glycosyltransferase that transfers N-acetylglucosamine residues on mannoses via a β1-6 linkage [39]. Elevated levels of β1-6 branched glycans are characteristic hallmarks of various tumors [40, 41], and overexpression of MGAT5 is frequently seen in cancer [42, 43]. As far as the molecular mechanism is concerned, Chiang et al. [44] found that MGAT5 catalyzed the N-glycosylation of CEACAM6 (carcinembryonic antigen-related cell adhesion molecule 6), a cell surface marker and interacting partner of epidermal growth factor receptor (EGFR). Glycosylation of CEACAM6 by MGAT5 is indispensable for EGFR clustering and activation as key signaling events leading to invasion and metastasis of oral squamous cell carcinoma.

CONCLUSIONS

Considering the above, we have identified one significantly associated SNP with lung cancer. Further investigations are needed on larger populations to boost the statistical power of this potentially novel glycobiomarker in cancer diagnostics. Our results suggest that the rs34944508 SNP might modulate the risk for lung cancer by influencing the expression of MGAT5, a typical cancer-associated glycosyltransferase. Nevertheless, further functional studies (such as luciferase assay) are required to test whether hsa-miR-3200-3p miRNA can indeed down-regulate MGAT5 expression levels in a sequence-specific manner.

FUNDING

This project was supported by the Hungarian National Research, Development and Innovation Office (NKFIH, OTKA) Grant K116263.

ACKNOWLEDGMENT
We would like to thank SCIEX for the kind support. Here we state that authors have no conflict of interest to declare.

REFERNCES

1. Jayaprakash NG, Surolia A. Role of glycosylation in nucleating protein folding and stability. The Biochemical journal. 2017;474(14):2333-47.
2. Peixoto A, Relvas-Santos M, Azevedo R, Santos LL, Ferreira JA. Protein Glycosylation and Tumor Microenvironment Alterations Driving Cancer Hallmarks. Frontiers in oncology. 2019;9:380.
3. Bassaganas S, Allende H, Cobler L, Ortiz MR, Llop E, de Bolos C, et al. Inflammatory cytokines regulate the expression of glycosyltransferases involved in the biosynthesis of tumor-associated sialylated glycans in pancreatic cancer cell lines. Cytokine. 2015;75(1):197-206.
4. Zoldos V, Novokmet M, Beceheli I, Lauc G. Genomics and epigenomics of the human glycome. Glycoconjugate journal. 2013;30(1):41-50.
5. B SG, Mohan Reddy P, Kottekad S. Comparative Site-Specific N-Glycosylation Analysis of Lactoperoxidase from Buffalo and Goat Milk Using RP-UHPLC-MS/MS Reveals a Distinct Glycan Pattern. Journal of agricultural and food chemistry. 2018;66(43):11492-9.
6. Lageveen-Kammeijer GSM, de Haan N, Mohaupt P, Wagt S, Filius M, Nouta J, et al. Highly sensitive CE-ESI-MS analysis of N-glycans from complex biological samples. Nature communications. 2019;10(1):2137.
7. Goulabchand R, Vincent T, Batteux F, Eliaou JF, Guilpain P. Impact of autoantibody glycosylation in autoimmune diseases. Autoimmunity reviews. 2014;13(7):742-50.
8. Huang C, Zhan T, Liu Y, Li Q, Wu H, Ji D, et al. Glycomic profiling of carcinoembryonic antigen isolated from human tumor tissue. Clinical proteomics. 2015;12(1):17.
9. Kovacs Z, Simon A, Szabo Z, Nagy Z, Varoczy L, Pal I, et al. Capillary electrophoresis analysis of N-glycosylation changes of serum paraproteins in multiple myeloma. Electrophoresis. 2017;38(17):2115-23.
10. Tozawa-Ono A, Kubota M, Honma C, Nakagawa Y, Yokomichi N, Yoshioka N, et al. Glycan profiling using formalin-fixed, paraffin-embedded tissues: Hippeastrum hybrid lectin is a sensitive biomarker for squamous cell carcinoma of the uterine cervix. The journal of obstetrics and gynaecology research. 2017;43(8):1326-34.
11. Ceciliani F, Pocacqua V. The acute phase protein alpha1-acid glycoprotein: a model for altered glycosylation during diseases. Current protein & peptide science. 2007;8(1):91-108.
12. Novokmet M, Lukic E, Vuckovic F, Ethuric Z, Keser T, Rajsl K, et al. Changes in IgG and total plasma protein glycomes in acute systemic inflammation. Scientific reports. 2014;4:4347.
13. Theodoratou E, Campbell H, Ventham NT, Kolarich D, Pucic-Bakovic M, Zoldos V, et al. The role of glycosylation in IBD. Nature reviews Gastroenterology & hepatology. 2014;11(10):588-600.
14. Drabik A, Bodzon-Kulakowska A, Suder P, Silberring J, Kulig J, Sierzega M. Glycosylation Changes in Serum Proteins Identify Patients with Pancreatic Cancer. Journal of proteome research. 2017;16(4):1436-44.
15. Krishnan S, Whitwell HJ, Cuenco J, Gentry-Maharaj A, Menon U, Pereira SP, et al. Evidence of Altered Glycosylation of Serum Proteins Prior to Pancreatic Cancer Diagnosis. International journal of molecular sciences. 2017;18(12).

16. Choi JW, Moon BI, Lee JW, Kim HJ, Jin Y. Use of CA153 for screening breast cancer: An antibodylectin sandwich assay for detecting glycosylation of CA153 in sera. Oncology reports. 2018;40(1):145-54.
17. Kawaguchi-Sakita N, Kaneshiro-Nakagawa K, Kawashima M, Sugimoto M, Tokiwa M, Suzuki E, et al. Serum immunoglobulin G Fc region N-glycosylation profiling by matrix-assisted laser desorption/ionization mass spectrometry can distinguish breast cancer patients from cancer-free controls. Biochemical and biophysical research communications. 2016;469(4):1140-5.
18. Ruhaak LR, Kim K, Stroble C, Taylor SL, Hong Q, Miyamoto S, et al. Protein-Specific Differential Glycosylation of Immunoglobulins in Serum of Ovarian Cancer Patients. Journal of proteome research. 2016;15(3):1002-10.
19. Weiz S, Wieczorek M, Schwedler C, Kaup M, Braicu EI, Sehouli J, et al. Acute-phase glycoprotein N-glycome of ovarian cancer patients analyzed by CE-LIF. Electrophoresis. 2016;37(11):1461-7.
20. Kazuno S, Fujimura T, Arai T, Ueno T, Nagao K, Fujime M, et al. Multi-sequential surface plasmon resonance analysis of haptoglobin-lectin complex in sera of patients with malignant and benign prostate diseases. Analytical biochemistry. 2011;419(2):241-9.
21. Ju L, Wang Y, Xie Q, Xu X, Li Y, Chen Z. Elevated level of serum glycoprotein bifucosylation and prognostic value in Chinese breast cancer. Glycobiology. 2016;26(5):460-71.
22. Mondal G, Saroha A, Bose PP, Chatterjee BP. Altered glycosylation, expression of serum haptoglobin and alpha-1-antitrypsin in chronic hepatitis C, hepatitis C induced liver cirrhosis and hepatocellular carcinoma patients. Glycoconjugate journal. 2016;33(2):209-18.
23. de Andrade M, Li Y, Marks RS, Deschamps C, Scanlon PD, Olswold CL, et al. Genetic variants associated with the risk of chronic obstructive pulmonary disease with and without lung cancer. Cancer Prev Res (Phila). 2012;5(3):365-73.
24. Ayyub A, Saleem M, Fatima I, Tariq A, Hashmi N, Musharraf SG. Glycosylated Alpha-1-acid glycoprotein 1 as a potential lung cancer serum biomarker. The international journal of biochemistry & cell biology. 2016;70:68-75.
25. Liang Y, Ma T, Thakur A, Yu H, Gao L, Shi P, et al. Differentially expressed glycosylated patterns of alpha-1-antitrypsin as serum biomarkers for the diagnosis of lung cancer. Glycobiology. 2015;25(3):331-40.
26. Togayachi A, Iwaki J, Kaji H, Matsuzaki H, Kuno A, Hirao Y, et al. Glycobiomarker, Fucosylated Short-Form Secretogranin III Levels Are Increased in Serum of Patients with Small Cell Lung Carcinoma. Journal of proteome research. 2017;16(12):4495-505.
27. Varadi C, Mittermayr S, Szekrenyes A, Kadas J, Takacs L, Kurucz I, et al. Analysis of haptoglobin N-glycome alterations in inflammatory and malignant lung diseases by capillary electrophoresis. Electrophoresis. 2013;34(16):2287-94.
28. Phelan CM, Tsai YY, Goode EL, Vierkant RA, Fridley BL, Beesley J, et al. Polymorphism in the GALNT1 gene and epithelial ovarian cancer in non-Hispanic white women: the Ovarian Cancer Association Consortium. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2010;19(2):600-4.

29. Kovacs-Nagy R, Elek Z, Szekely A, Nanasi T, Sasvari-Szekely M, Ronai Z. Association of aggression with a novel microRNA binding site polymorphism in the wolframin gene. American journal of medical genetics Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics. 2013;162B(4):404-12.
30. Elek Z, Nemeth N, Nagy G, Nemeth H, Somogyi A, Hosszufalusi N, et al. Micro-RNA Binding Site Polymorphisms in the WFS1 Gene Are Risk Factors of Diabetes Mellitus. PloS one. 2015;10(10):e0139519.
31. Srinivasan S, Stephens C, Wilson E, Panchadsaram J, DeVoss K, Koistinen H, et al. Prostate Cancer Risk-Associated Single-Nucleotide Polymorphism Affects Prostate-Specific Antigen Glycosylation and Its Function. Clinical chemistry. 2019;65(1):e1-e9.
32. Vajaria BN, Patel PS. Glycosylation: a hallmark of cancer? Glycoconjugate journal. 2017;34(2):147-56.
33. Banlaki Z, Elek Z, Nanasi T, Szekely A, Nemoda Z, Sasvari-Szekely M, et al. Polymorphism in the serotonin receptor 2a (HTR2A) gene as possible predisposal factor for aggressive traits. PloS one. 2015;10(2):e0117792.
34. Elek Z, Denes R, Prokop S, Somogyi A, Yowanto H, Luo J, et al. Multicapillary gel electrophoresis based analysis of genetic variants in the WFS1 gene. Electrophoresis. 2016;37(17-18):2313-21.
35. Zhou F, Wang W, Xing Y, Wang T, Xu X, Wang J. NF-kappaB target microRNAs and their target genes in TNFalpha-stimulated HeLa cells. Biochimica et biophysica acta. 2014;1839(4):344-54.
36. Satoh J, Kino Y, Niida S. MicroRNA-Seq Data Analysis Pipeline to Identify Blood Biomarkers for Alzheimer's Disease from Public Data. Biomarker insights. 2015;10:21-31.
37. Groen K, Maltby VE, Lea RA, Sanders KA, Fink JL, Scott RJ, et al. Erythrocyte microRNA sequencing reveals differential expression in relapsing-remitting multiple sclerosis. BMC medical genomics. 2018;11(1):48.
38. Zhou X, Chen H, Wang Q, Zhang L, Zhao J. Knockdown of Mgat5 inhibits CD133+ human pulmonary adenocarcinoma cell growth in vitro and in vivo. Clinical and investigative medicine Medecine clinique et experimentale. 2011;34(3):E155-62.
39. Nagae M, Kizuka Y, Mihara E, Kitago Y, Hanashima S, Ito Y, et al. Structure and mechanism of cancer-associated N-acetylglucosaminyltransferase-V. Nature communications. 2018;9(1):3380.
40. Huang W, Luo WJ, Zhu P, Tang J, Yu XL, Cui HY, et al. Modulation of CD147-induced matrix metalloproteinase activity: role of CD147 N-glycosylation. The Biochemical journal. 2013;449(2):437-48.
41. Ferdosi S, Rehder DS, Maranian P, Castle EP, Ho TH, Pass HI, et al. Stage Dependence, Cell-Origin Independence, and Prognostic Capacity of Serum Glycan Fucosylation, beta1-4 Branching, beta1-6 Branching, and alpha2-6 Sialylation in Cancer. Journal of proteome research. 2018;17(1):543-58.
42. Bubka M, Link-Lenczowski P, Janik M, Pochec E, Litynska A. Overexpression of N-acetylglucosaminyltransferases III and V in human melanoma cells. Implications for MCAM N-glycosylation. Biochimie. 2014;103:37-49.

43. Liu J, Liu H, Zhang W, Wu Q, Liu W, Liu Y, et al. N-acetylglucosaminyltransferase V confers hepatoma cells with resistance to anoikis through EGFR/PAK1 activation. Glycobiology. 2013;23(9):1097-109.
44. Chiang WF, Cheng TM, Chang CC, Pan SH, Changou CA, Chang TH, et al. Carcinoembryonic antigen-related cell adhesion molecule 6 (CEACAM6) promotes EGF receptor signaling of oral squamous cell carcinoma metastasis via the complex N-glycosylation. Oncogene. 2018;37(1):116-27.

TABLES

Table 1. Primers used in the single base extension method in combination with capillary gel electrophoresis. The two primers designated with "PCR:" were used in the first step to amplify the flanking region of the investigated SNP, whereas the primer labeled with "SBE:" was applied in the single base extension reaction.

SNP	Primers (uppers: amplification, lower: SBE reaction)	T_m (°C)	Product
rs2922471 C/T	PCR: 5' TGGGGCTTTCAGAAAAGGCCTTGCCCA 3'	62.8	548 bp
	PCR: 5' GGAGCCCCAGGTCATCCAGCTGTTGC 3'	65.8	
	SBE: 5' GTTTTTCAGCTACATTAATATATTGT GAGACCACTGTCTTA 3'	61.5	42 nt
rs4736675 A/G	PCR: 5' GTGGCCTTCTCACACCCTCCCGTG 3'	64.2	268 bp
	PCR: 5' CCCTGCCACTGCCTGCCCAGAAC 3'	64.2	
	SBE: 5' CCAGGGATGAGTTCCAGTTGGTTTAA GCCA 3'	63.0	30 nt
rs1042757 C/G	PCR: 5' AGGAGGTGGCATTCCGACAGCAGG 3'	62.5	573 bp
	PCR: 5' GAGCCAGTGGGTAAGGACTGGGG 3'	62.4	
	SBE: 5' TCTGGTGAGATGTTTCATATTTGTGA CAGTTAATTTAAAAATTATGA 3'	61.1	48 nt

Table 2. Single nucleotide polymorphisms investigated by TaqMan™ OpenArray™ in the case–control association study. The major allele stands at the first position in the "*Alleles*" column. Chromosomal localization is given according to genome build GRCh38. Statistical *p* value of the χ^2-analysis of the Hardy–Weinberg equilibrium is shown. *MAF*: Minor allele frequency in the whole sample.

	dbSNP	*Alleles*	*Chr.*	*Position*	*Gene*	*Call rate*	*Hardy–Weinberg*	*MAF*
1	rs2230908	CA	2	134 448 769	MGAT5	0.0%	n/a	n/a
2	rs34944508	CT	2	134 450 393	MGAT5	97.2%	0.818	0.053
3	rs79594066	TC	2	134 452 524	MGAT5	98.0%	0.987	0.046
4	rs35166820	CT	3	186 928 351	ST6GAL1	95.9%	1.000	0.112
5	rs9814673	CT	3	186 929 863	ST6GAL1	83.5%	0.003	0.329
6	rs28366038	GA	3	186 929 991	ST6GAL1	92.6%	0.872	0.062
7	rs1468906	AG	3	186 930 328	ST6GAL1	47.4%	1.000	0.099
8	rs1042642	AG	3	187 075 641	ST6GAL1	97.4%	0.499	0.472
9	rs2239611	GA	3	187 075 914	ST6GAL1	96.3%	1.000	0.014
10	rs2284750	CT	3	187 076 303	ST6GAL1	93.5%	1.000	0.038
11	rs1801380	GA	3	187 077 852	ST6GAL1	88.5%	0.555	0.138
12	rs7559	TC	3	187 078 265	ST6GAL1	94.2%	1.000	0.145
13	rs1042757	GC	3	187 078 342	ST6GAL1	99.2%	0.429	0.474
14	rs4736674	AT	8	133 456 465	ST3GAL1	98.8%	1.000	0.186
15	rs4736675	GA	8	133 456 621	ST3GAL1	99.3%	1.000	0.186
16	rs16904924	TA	8	133 457 034	ST3GAL1	98.7%	1.000	0.186
17	rs11782689	TC	8	133 457 293	ST3GAL1	99.3%	0.852	0.178
18	rs2142306	TC	8	133 458 388	ST3GAL1	98.4%	1.000	0.448
19	rs2922467	GC	8	133 572 946	ST3GAL1	97.2%	0.835	0.256
20	rs2922471	CT	8	133 574 220	ST3GAL1	96.7%	1.000	0.449
21	rs3894326	AT	19	5 843 773	FUT3	98.5%	0.712	0.093
22	rs778986	GA	19	5 844 526	FUT3	98.4%	0.718	0.199
23	rs812936	AG	19	5 844 638	FUT3	98.9%	1.000	0.208
24	rs11673407	AG	19	5 851 325	FUT3	96.6%	1.000	0.381
25	rs2306969	GA	19	5 851 790	FUT3	97.5%	0.624	0.267
26	rs418821	CG	19	48 696 547	FUT2	98.4%	1.000	0.102
27	rs16982241	GA	19	48 699 602	FUT2	89.8%	0.313	0.119
28	rs28362834	CG	19	48 702 748	FUT2	97.5%	0.351	0.100
29	rs679574	CG	19	48 702 851	FUT2	99.2%	0.400	0.421
30	rs601338	GA	19	48 703 417	FUT2	99.1%	0.019	0.450
31	rs602662	GA	19	48 703 728	FUT2	98.5%	0.518	0.437
32	rs632111	AG	19	48 705 721	FUT2	98.6%	0.757	0.433

Table 3. Comparison of the two techniques used for SNP genotype determination

	TaqMan® OpenArray®	Capillary gel electrophoresis
throughput	medium	low
number of steps	1	3
format	stipulated	flexible
pre-designed assays	yes	no
number of samples	960	flexible
number of loci	12, 24, 60, 120, 180, 240	flexible
required DNA amount	~ 4 ng DNA / locus	< 1 ng DNA / locus
analysis of repeat variants	no	yes
possibility of repeating failed analyses	limited	yes

Table 4. Association analysis by case–control approach. Minor allele frequency values in the 4 subject groups, nominal *p* values of each individual χ^2-test as well as the statistical *p* value after correction for multiple testing by Benjamini–Hochberg procedure (false discovery rate, FDR) are shown.

SNP	Minor allele frequency (MAF)				nominal	FDR
	control	tumor	COPD	comorbid	*p*	*p*
rs632111	0.463	0.442	0.403	0.417	0.2284	0.4283
rs418821	0.110	0.101	0.106	0.076	0.5179	0.6474
rs601338	0.477	0.459	0.426	0.429	0.3312	0.4969
rs602662	0.467	0.450	0.403	0.416	0.1552	0.4233
rs3894326	0.091	0.082	0.115	0.075	0.1969	0.4220
rs778986	0.180	0.200	0.209	0.214	0.5901	0.7082
rs28362834	0.094	0.087	0.110	0.120	0.4336	0.5913
rs812936	0.190	0.202	0.221	0.227	0.5103	0.6656
rs11673407	0.396	0.390	0.365	0.367	0.6930	0.7700
rs2306969	0.260	0.239	0.310	0.254	0.0587	0.5866
rs34944508	0.087	0.033	0.045	0.041	**0.0007**	**0.0207**
rs79594066	0.053	0.060	0.032	0.041	0.1469	0.4407
rs2922471	0.438	0.441	0.457	0.475	0.7592	0.8135
rs4736675	0.170	0.203	0.173	0.213	0.2946	0.4910
rs35166820	0.122	0.087	0.103	0.168	**0.0085**	0.1268
rs11782689	0.165	0.193	0.164	0.207	0.3300	0.5211
rs16904924	0.171	0.202	0.173	0.211	0.3503	0.5005
rs2142306	0.429	0.485	0.429	0.450	0.2168	0.4337
rs4736674	0.170	0.205	0.172	0.211	0.2918	0.5150
rs679574	0.458	0.427	0.393	0.392	0.1316	0.4386
rs1042757	0.466	0.448	0.478	0.455	0.0854	0.6404
rs1042642	0.465	0.447	0.472	0.458	0.1071	0.4017
rs2922467	0.261	0.248	0.270	0.232	0.6851	0.7905
rs1801380	0.120	0.121	0.151	0.179	0.0887	0.5322
rs2239611	0.015	0.018	0.011	0.017	0.8091	0.8370
rs16982241	0.127	0.132	0.093	0.135	0.1789	0.4129
rs2284750	0.023	0.043	0.048	0.033	0.1723	0.4308
rs7559	0.124	0.132	0.158	0.186	0.0921	0.4603
rs28366038	0.056	0.069	0.062	0.063	0.8817	0.8817
rs9814673	0.360	0.291	0.318	0.367	0.1040	0.4458

FIGURES

Figure 1. SNP genotyping by TaqMan™ probes and by capillary gel electrophoresis. Analysis of the rs4736675 SNP is shown. *A*: The two allele-specific probes are labeled with different fluorophores (FAM and VIC) in case of the real-time PCR based method. Thus, samples are separated into 3 clusters based on the ratio of the two fluorescent signals according to their genotype. *B*: The primer extension method employs a primer annealing to the nucleotide nearby the locus of interest. This is elongated by a fluorescently labeled, chain terminator acyclo nucleotide. The obtained products were separated and detected by capillary gel electrophoresis.

Figure 2. Allele-wise association analysis of the rs34944508 SNP. The diagram shows the allele-frequency values of the SNP in the four subject groups. It is notable that there is >2× difference between the minor allele frequencies of controls and lung cancer subjects.