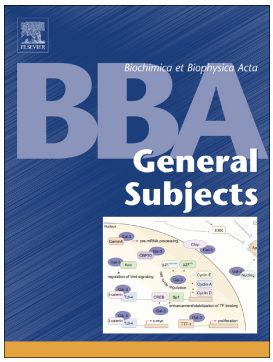
IgG glycosylation and DNA methylation are interconnected with smoking



Annika Wahl, Silva Kasela, Elena Carnero Monotoro, Maarten van Iterson, Jerko Štambuk, Sapna Sharma, Erik van den Akker, Lucija Klaric, Elisa Benedetti, Genadij Razdorov, Irena Trbojević-Akmačić, Frano Vučković, Ivo Ugrina, Marian Beekman, Joris Deelen, Diana van Heemst, Bastiaan T. Heijmans, B.I.O.S. Consortium, Manfred Wuhrer, Rosina Plomp, Toma Keser, Mirna Šimurina, Tamara Pavić, Ivan Gudelj, Jasminka Krištić, Harald Grallert, Sonja Kunze, Annette Peters, Jordana T. Bell, Timothy D. Spector, Lili Milani, P. Eline Slagboom, Gordan Lauc, Christian Gieger

PII: S0304-4165(17)30341-0
DOI: doi:10.1016/j.bbagen.2017.10.012
Reference: BBAGEN 28966

To appear in:

Received date: 19 June 2017
Revised date: 1 October 2017
Accepted date: 16 October 2017

Please cite this article as: Annika Wahl, Silva Kasela, Elena Carnero Monotoro, Maarten van Iterson, Jerko Štambuk, Sapna Sharma, Erik van den Akker, Lucija Klaric, Elisa Benedetti, Genadij Razdorov, Irena Trbojević-Akmačić, Frano Vučković, Ivo Ugrina, Marian Beekman, Joris Deelen, Diana van Heemst, Bastiaan T. Heijmans, B.I.O.S. Consortium, Manfred Wuhrer, Rosina Plomp, Toma Keser, Mirna Šimurina, Tamara Pavić, Ivan Gudelj, Jasminka Krištić, Harald Grallert, Sonja Kunze, Annette Peters, Jordana T. Bell, Timothy D. Spector, Lili Milani, P. Eline Slagboom, Gordan Lauc, Christian Gieger , IgG glycosylation and DNA methylation are interconnected with smoking. The address for the corresponding author was captured as affiliation for all authors. Please check if appropriate. Bbagen(2017), doi:10.1016/j.bbagen.2017.10.012

be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

IgG Glycosylation and DNA methylation are interconnected with smoking

Authors

Annika Wahl[1,2], Silva Kasela[3], Elena Carnero Monotoro[4], Maarten van Iterson[5], Jerko Štambuk[6], Sapna Sharma[1,2], Erik van den Akker[5,7], Lucija Klaric[6,8,9], Elisa Benedetti[10], Genadij Razdorov[6], Irena Trbojević-Akmačić[6], Frano Vučković[6], Ivo Ugrina[11,12], Marian Beekman[5], Joris Deelen[5, 14], Diana van Heemst[15], Bastiaan T. Heijmans[5], BIOS Consortium[16], Manfred Wuhrer[13], Rosina Plomp[13], Toma Keser[11], Mirna Šimurina[11], Tamara Pavić[11], Ivan Gudelj[6], Jasminka Krištić[6], Harald Grallert[1,2,17], Sonja Kunze[1,2], Annette Peters[2], Jordana T. Bell[4] , Timothy D. Spector[4], Lili Milani[3], P. Eline Slagboom[5], Gordan Lauc[6,11], Christian Gieger[1,2]

Affiliations

[1]Research Unit Molecular Epidemiology, Institute of Epidemiology 2, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany

[2]Institute of Epidemiology 2, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany

[3]Estonian Genome Center, University of Tartu, Tartu, Estonia

[4]Department of Twin Research and Genetic Epidemiology, King's College London, London, UK

[5] Department of Molecular Epidemiology, Leiden University Medical Center (LUMC), Leiden, The Netherlands

[6]Genos Glycoscience Research Laboratory, Zagreb, Croatia

[7] Pattern Recognition & Bioinformatics, Delft University of Technology, Delft, The Netherlands

[8]Centre for Population Health Sciences, School of Molecular, Genetic and Population Health Sciences, University of Edinburgh, Edinburgh, United Kingdom

[9]MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, United Kingdom

[10]Institute of Computational Biology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany

[11]University of Zagreb, Faculty of Pharmacy and Biochemistry, Zagreb, Croatia

[12]University of Split, Faculty of Science, Split, Croatia

[13] Center for Proteomics & Metabolomics, Leiden University Medical Center (LUMC), Leiden, The Netherlands

[14]Max Planck Institute for Biology of Ageing, Köln, Germany

[15] Department of Internal Medicine, Section Gerontology and Geriatrics, Leiden University Medical Center (LUMC), Leiden, The Netherlands

[16] The Biobank-Based Integrative Omics Study (BIOS) Consortium (supplemental material S6 provides a full list of the members of the consortium)

[17] German Center for Diabetes Research (DZD), Neuherberg, Germany

Highlights

- An epigenome-wide association study including 3000 individuals from four cohorts was performed
- DNA methylation within 5 genomic locations is associated with IgG glycosylation
- DNA methylation mediated the effects of smoking on IgG glycosylation

Abstract

Background: Glycosylation is one of the most common post-translation modifications with large influences on protein structure and function. The effector function of immunoglobulin G (IgG) alters between pro- and anti-inflammatory, based on its glycosylation. IgG glycan synthesis is highly complex and dynamic.

Methods: With the use of two different analytical methods for assessing IgG glycosylation, we aim to elucidate the link between DNA methylation and glycosylation of IgG by means of epigenome-wide association studies. In total, 3,000 individuals from 4 cohorts were analyzed.

Results: The overlap of the results from the two glycan measurement panels yielded DNA methylation of 7 CpG-sites on 5 genomic locations to be associated with IgG glycosylation: cg25189904 (chr.1, *GNG12*); cg05951221, cg21566642 and cg01940273 (chr.2, *ALPPL2*); cg05575921 (chr.5, *AHRR*); cg06126421 (6p21.33); and cg03636183 (chr.19, *F2RL3*). Mediation analyses with respect to smoking revealed that the effect of smoking on IgG glycosylation may be at least partially mediated via DNA methylation levels at these 7 CpG-sites.

Conclusion: Our results suggest the presence of an indirect link between DNA methylation and IgG glycosylation that may in part capture environmental exposures.

General Significance: An epigenome-wide analysis conducted in four population-based cohorts revealed an association between DNA methylation and IgG glycosylation patterns. Presumably, DNA methylation mediates the effect of smoking on IgG glycosylation.

Keywords

DNA methylation; IgG glycosylation; smoking; EWAS; mediation

Introduction

Functional diversity of the proteome is mainly achieved by post-translational modifications. One of the most common modifications is glycosylation, which affects protein structure and function. A well suited glycoprotein model is immunoglobulin G (IgG), the most abundant immunoglobulin circulating in blood [1], as its functional importance and physiological significance has previously been described [1, 2]. Two identical heavy and light chains, respectively, form the IgG polypeptide backbone [3]. The hinge region connecting the antigen-binding part (Fab) with the crystallizable region (Fc) part of IgG is the most variable component and mainly differentiates the IgG subclasses (IgG1, IgG2, IgG3 and IgG4) [3].

All IgG molecules contain a conserved *N*-glycosylation site on the Fc portion that controls pro- and anti-inflammatory functions of the glycoprotein [4, 5]. Additionally, about 15 to 25% of IgG molecules contain *N*-linked glycans in the Fab domain with several functions of their own [6]. The composition of the attached glycans is itself influenced by several factors, such as enzymes and the availability of precursors. *N*-glycan remodeling happens in the Golgi and the endoplasmic reticulum (ER) [7] and underlies the interaction of donor molecules, co-factors and enzymes including glycosyltransferases and glycosidases [8].

The complex interaction needed to compose the final functional glycan structure therefore cannot be encoded by a single gene or a combination of genes [8]. Still, the outcome of the enzymatic activities is well regulated and several genes encoding glycosyltransferases have been shown to associate with differing IgG glycosylation patterns [9].

While genetic regulation of the glycosylation synthesis could be partially elucidated [10, 11], the impact of epigenetics on IgG glycosylation is still not sufficiently studied. However, it is widely accepted and supported by different studies that the environment and, consequently, epigenetics plays an important role for the *N*-glycome [8, 12-15]. Additionally, it has been reviewed in [14] and [15] that epigenetic mechanisms have a key function in the IgG glycosylation pathway.

DNA methylation is one of the best studied epigenetic regulations leading to differential gene expression [16-18], especially gene silencing due to heavy methylation [19, 20]. Associations between DNA methylation and several physical and environmental factors, including smoking [21], lipid levels [22, 23] and physical activity [24], have been established over the last years. Indeed, using Mendelian randomization, it was recently shown that an altered DNA methylation status results from variations of blood lipid levels [23], highlighting the direction of causation for specific blood lipid levels. IgG glycosylation compositions have been shown to be altered in several diseases [25], including cancer [26, 27] and to change with age and hypertension as well as with smoking and with

other environmental influences [28, 29]. Effects of DNA methylation on genes involved in IgG glycan synthesis [30] and on the global *N*-glycan profile have especially been found in cancer [31, 32].

By analyzing the relation between DNA methylation and IgG glycans, we aim to immerse deeper into the epigenetic regulation of IgG glycosylation.

In this study, we compare epigenome-wide analyses conducted in four cohorts, two cohorts for each IgG glycan panel including 609, 148, 619 and 1,630 samples, respectively. The glycans measured by ultra-performance liquid chromatography (UPLC) reflect IgG glycosylation from both Fab and Fc parts, while the measurements obtained by liquid chromatography electrospray mass spectrometry (LC/MS) describe the glycan structures at the Fc part of the IgG glycopeptides, separately for each subclass. We discuss the results separately for the two panels and combine the two approaches by analyzing their shared epigenetic associations. Furthermore, we conduct a mediation analysis for one cohort per glycan panel to delineate the relation of IgG glycans, DNA methylation and smoking.

With this study we hope to elucidate the role of epigenetics in the entangled pathway leading to definite regulatory IgG glycan compositions.

Methods

Figure 1 summarizes the initial analysis steps as well as the additional mediation analysis. In total, 3,200 samples were available for our analyses. We studied two different methods of IgG glycosylation, UPLC (left/blue side in Figure 1) and LC/MS (right/green side in Figure 1). UPLC data was available in TwinsUK (UK adult twin registry; n = 877) [33] and EGCUT (Estonian Genome Center, University of Tartu; n = 148) [34], while LC/MS was available in KORA F4 (Cooperative Health Research in the Augsburg Region; n = 1,630) [35] and LLS (Leiden Longevity Study; n = 619) [36]. The overlapping results from both IgG glycan panels are presented in the middle of Figure 1 and are followed by the results from the additional mediation analysis.

Study Cohorts

General Characteristics on the four cohorts are summarized in table S1.

Cooperative Health Research in the Augsburg Region (KORA F4)

The KORA F4 study, conducted in 2006-2008, is an independent population-based health survey [35]. The study followed the recommendations of the Declaration of Helsinki and was approved by the local ethical committees.

Information on sociodemographic variables including sex, age, BMI (in kg/m²) was collected by trained medical staff during a standardized interview. White blood cell count (WBC) was measured

from whole blood plasma and estimated white blood cell proportions were obtained using the method by Houseman et al. [37]. By definition the cell proportions sum up to one, so we excluded from the analysis the least common cell type (NK) with the lowest mean to avoid confounding.

In KORA F4, a total of 3,080 persons participated of whom 1,630 have both DNA methylation and IgG glycopeptides measured. For the mediation analysis in KORA F4, we had 1,594 samples with known smoking status including 671 non-smokers and 923 ever smokers (691 former smokers and 232 current smokers).

Leiden Longevity Study (LLS)

The study protocol of LLS was approved by the local medical ethical committee and the study was conducted according to the recommendations of the Declaration of Helsinki. Written informed consent was obtained from all study participants. LLS collected data on long-lived individuals, their offspring and the partners of the offspring, as controls [36]. For this current study offspring and controls persons have been analyzed (n=619, mean age 59.2 years (range 39-81 years)) with both DNA methylation and IgG glycosylation data available [38].

UK adult twin registry (TwinsUK)

The TwinsUK cohort includes more than 13,000 monozygotic and dizygotic twin volunteers from all regions across the United Kingdom. Volunteers were not recruited on the basis of any specific trait or disease and have been shown to have diseases and lifestyle characteristics similar to the general population [33, 39] cohort. DNA methylation was measured for 877 randomly selected individuals, from which 609 had information on glycan measurements. White blood cell proportions were estimated using the method by Houseman et al. [37].

Estonian Genome Center, University of Tartu (EGCUT)

The Estonian Genome Center at the University of Tartu (EGCUT) [34] is a population based biobank which comprises health, genealogical and 'omics' data of close to 52,000 individuals ≥ 18 years of age, closely reflecting the age distribution in the adult Estonian population [34]. The collection of blood samples and the data is conducted according to the Estonian Human Gene Research Act and all participants have signed a broad informed consent.

Participants of the EGCUT have been recruited by clinicians at their offices or data collectors at recruitment offices of the EGCUT. A computer assisted personal interview was completed for each participant, including personal data (place of birth, place(s) of living, nationality etc.), genealogical data (family history, three generations), educational and occupational history and lifestyle data (physical activity, dietary habits, smoking, alcohol consumption, quality of life).

A blood sample was drawn from each participant during the recruitment. Estimated white blood cell proportions were obtained using the method by Houseman et al. [37].

In total, 148 individuals have both DNA methylation and IgG glycan measurements. They were selected from a subgroup of "healthy" individuals who have been re-contacted for a second time-point interview, a subgroup of individuals that died of cancer within five years since recruitment („healthy" at baseline), and a subgroup of men diagnosed with prostate cancer within five years since recruitment („healthy" at baseline). Of them, 45 are current, 39 former and 64 never-smokers.

DNA isolation and preparation

A blood sample was drawn from each participant during the recruitment and DNA was isolated using standard procedures. DNA methylation was analyzed with the Illumina HumanMethylation450K BeadChip [40]. Preceding treatment of the blood samples followed the manufacturer's guidelines. DNA methylation data were measured with the Illumina HumanMethylation450K BeadChip [40] in all four cohorts.

DNA Methylation Preprocessing

Methylation data from all studies underwent specific initial quality control. Beta-values of DNA methylation as well as appropriate technical covariates were included in the statistical models. Family structure was accounted for if needed.

For KORA F4, processing of the raw DNA methylation data and initial quality control was accomplished as described in detail in [21]. Following, data from KORA F4 and EGCUT followed the quantile normalization pipeline described by Lehne et al [41]. The pipeline includes a background correction, a detection *p*-value threshold of 0.01 and a sample callrate of 95%. Quantile normalization of the intensity values was performed and beta-values for methylation were used for further analyses. Markers on both X and Y chromosomes had been excluded at the time of the analysis. To avoid technical confounding a principle component analysis (PCA) was performed for the positive control probes of the BeadChip. Both the first 20 control probe PCAs as well as white blood cell composition proportions were calculated and used in the final model. No callrate threshold for CpG-sites was used.

Data from TwinsUK underwent an initial quality control including e.g. a detection p-value > 0.05 and the exclusion of DNA methylation probes that mapped incorrectly or to multiple locations. Following, the data were beta-mixture quantile normalized as described in [42].

For LLS, the generation of genome-wide DNA methylation data is described by Bonder et al [43]. The original idat files were generated by the Illumina iScan BeadChip scanner. Subsequently, sample

quality control was performed using MethylAid [44]. Similar to the pipeline used for the other cohorts, probes with a high detection *p*-value (> 0.01), probes with a low bead count (< 3 beads), and probes with a low success rate (missings in > 95 %) were excluded from the analysis. Subsequently, imputation was performed to impute the missing values. Functional normalization, as implemented in the minfi package, was used.

Measurement of IgG Glycosylation

The two different methods applied, UPLC and LC/MS as well as the single process steps applied have been described in detail in [9, 11, 45-47]. In short, for both analytical methods, IgG was first isolated from plasma as presented in [9].

For KORA F4 and LLS, glycosylation patterns were examined by the analysis of tryptic glycopeptides using the nanoLC-ESI-MS (denoted as LC/MS) method as outlined in [46] which provides subclass-specific glycoforms. Due to their similar peptide moieties however, IgG2 and IgG3 cannot be distinguished in Caucasian [48] and will be denoted by IgG2/3 or simply IgG2 in the following. For LLS, IgG was isolated by protein A, which does not bind to IgG3, so IgG2 data represents pure IgG2 subclass. More details on the analytical method can be found in [47]. In total, 20 glycoforms could be observed for both IgG1 and IgG2/3, and 10 peaks describing fucoslyated glycoforms for IgG4.

Glycan profiles for TwinsUK and EGCUT, were investigated by UPLC after IgG isolation and IgG *N*-glycan release and labeling. Details are provided in [9, 45, 47]. In total, 24 glycan peaks can be distinguished by this approach. For TwinsUK however, the peaks of GP20 and GP21 (see table S2a for details) were not uniquely distinguishable and both traits are excluded in the downstream analysis.

Summarizing, LC/MS measurements provide data on the glycan structure attached to the Fc part of IgG glycopeptides. UPLC data in contrast contain information on the total released IgG glycans from both Fab and Fc parts.

IgG Glycan Preprocessing

Both UPLC and LC/MS measured IgG *N*-glycans were total area normalized as implemented in the R-package 'glycanr' v0.3.0 [49]. We consider glycan percentages for UPLC and glycan percentages per subclass for LC/MS data. Normalization was followed by log-transformation and batch correction (per subclass for LC/MS data) using the ComBat algorithm [50] of the R package 'sva' v3.14.0 [51]. In order to recapture the initial scale, the data were exponentiated and then derived traits were calculated as provided by the glycanr-package [49] and as in [45]. See the tables S2a and S2b for an overview of initial and derived traits for both UPLC and LC/MS data. To ensure normal distribution of

the phenotypes, data were inverse normal rank transformed [52]. For similar derived traits however, the normal rank transformation led to the exactly same values for differing traits (e.g. LC_IGP42 and LC_IGP44; LC_IGP43 and LC_IGP45 in the KORA F4 cohort). We kept those similar traits and flagged them within each cohort separately.

Statistical Methods

Primary EWAS analysis

In each cohort we performed linear regression models for all glycan traits with each glycan as outcome. Beta-values of DNA methylation entered the model as independent variables. To account for confounding, we adjusted the model by age, sex, BMI, white blood cell count (WBC) (available only in KORA F4), estimated white blood cell proportions and technical covariates if applicable. Different software was used for the initial computation of the linear model (KORA F4: omicABEL [53], LLS: R-package cate v1.0.4 [54], TwinsUK and EGCUT: R (3.2.2) [55]).

Meta-analysis

For UPLC and LC/MS data respectively, we performed a random effects meta-analysis using the package 'metafor' [56] from the statistical software R (Version 3.2.1). Inverse variance weighting and the restricted maximum likelihood estimator were used for the analysis.

We excluded six glycan traits (LC_IGP170, LC_IGP148, LC_IGP147, LC_IGP149, LC_IGP107, LC_IGP150) in the LC/MS dataset due to their deviating QQplots (an example is shown in supplemental figure 1) in the meta-analysis in order to reduce bias from measurement errors. Additionally, these traits showed an uncommonly large number of associations (> 5,000) randomly across the genome.

We set the Bonferroni-corrected [57] epigenome-wide significant *p*-value threshold for UPLC data at 4.6×10^{-9} (0.05/ 450,000 CpG-sites / 24 initially measured glycans) and at 2.2×10^{-9} for LC/MS data (0.05/ 450,000 CpG-sites / 50 initially measured glycans). As suggestive threshold for both panels, we used 1.1×10^{-7} (0.05/ 450,000 CpG-sites). In order to assure stability of the results, we additionally excluded 248 combinations with differing effect directions and *p*-values for an individual study larger than 0.05 from the meta-analysis results (between 43 IgG glycan traits and 137 CpG-sites). A full list of all associations before exclusion can be found in table S4b.

To combine the results from both meta-analyses, we took all suggestive CpG-sites ($p < 1.1 \times 10^{-7}$) from both panels and analyzed their associations with IgG glycan traits with *p*-values below 0.001 in the other panel, respectively. For further analyses, we focused on CpG-sites that were suggestive in one meta-analysis and had a maximal *p*-value of 0.001 for any association within the other glycan panel.

Mediation analysis

CpG-sites and their associated IgG glycans remaining after the exclusion criteria were further used in a mediation analysis with respect to smoking. We performed Sobel tests [58, 59] to assess the mediation effect and performed an analysis of variance for the appropriate CpG - glycan combinations. All analyses were conducted in R by using the package 'bda' v5.1.6 [60] and the functions sobel.test and aov ('stats' v3.4.0 [55]). We differentiated between ever smokers (current or former smokers) and never smokers. While smoking status was considered as independent variable, the CpG-site entered the model either as mediator or as dependent variable whereas the glycan trait fulfilled the remaining task, correspondingly. With these two approaches we want to examine the direction of the mediation. The *p*-value thresholds for the UPLC and LC/MS dataset follow the Bonferroni-correction [57] and were set to 0.017 (0.05/ 3 combinations) and 6.8×10^{-4} (0.05/ 74 combinations), respectively (see Table 2a and 2b for additional information on the selected combinations).

Results

Figure 1 displays the analysis workflow for the identification of associations between CpG-sites and IgG glycosylation traits.

For the UPLC meta-analysis, we found in total 124 associations between 27 glycan traits and 26 CpG-sites to be at least suggestively significant at Bonferroni-correction ($p < 1.1 \times 10^{-7}$, Bonferroni-corrected for number of CpG-sites). Out of these associations, 61 were epigenome-wide significant ($p < 4.6 \times 10^{-9}$, Bonferroni-corrected for the number of CpG-sites and the number of initial measured UPLC glycan traits) associations corresponding with 7 CpG-sites and 14 glycan traits (Figure 1, table S3).

For the LC/MS data, we found in total 6 suggestive ($p < 1.1 \times 10^{-7}$, Bonferroni-corrected for number of CpG-sites) associations between 4 CpG-sites and 3 glycan traits (Figure 1, table S4a). Of these associations, 2 were epigenome-wide significance threshold of $p < 2.2 \times 10^{-9}$ (Bonferroni-corrected for the number of CpG-sites and the number of initial measured LC/MS glycopeptides), i.e. cg0612642: p(LC_IGP43) = 8.82×10^{-13}, p(LC_IGP45) = 1.01×10^{-12}.

Combining the results from the 26 CpG-sites from the UPLC glycan panel and the 4 CpG sites from the LC/MS glycan panel yields 28 unique CpG-sites, out of which 7 CpG-sites are suggestive ($p < 1.1 \times 10^{-7}$) for one glycan panel and have at least one association with $p < 0.001$ in the other glycan panel.

Of the 7 CpG sites (Figure 1, table T1), 6 CpG sites associated with 20 UPLC glycan trait and 3 CpG sites with 2 LC/MS IgG glycan traits (Table S5a and S5b). Two CpG-sites are suggestively significant for both glycan panels, cg05951221 and cg21566642 (both on chromosome 2).

The 7 CpG-sites are located on 5 genomic locations: cg25189904 on chromosome 1 (1:68,299,493; 1p31.3, *GNG12*); cg05951221, cg21566642, cg01940273 on chromosome 2 (2:233,284,402 – 2: 233,284,934; 2q37.1, close to *ALPPL2 [61]*) from which the first two are suggestive for both glycan panels; cg05575921 on chromosome 5 (5: 373,378; 5p15.33; *AHRR*); cg06126421 on chromosome 6 (6:30,720,080; 6p21.33); and cg03636183 on chromosome 19 (19: 17,000,585; 19p13.11, *F2RL3*).

The 3 CpG-sites on chromosome 2 are in very close proximity and belong to the same CpG Island located at chr2:233,283,398-233,285,959. Thus, we assume only one independent signal on chromosome 2.

For the UPLC IgG traits, 20 IgG glycan traits exhibit similar association to the CpG-sites and they mostly represent the percentage of fucosylation with and without bisecting GlcNAc, as well as the ratio between structures containing a bisecting GlcNAc and lacking a bisecting GlcNAc.

The associated IgG glycans in the LC/MS panel are 2 derived traits, LC_IGP43 and LC_IGP45, which can be described them as the relation of fucosylated monosialylated structures with bisecting GlcNAc to all fucosylated and monosialyated structures within IgG1.

In general, a decrease of DNA methylation at the CpG-sites on chromosome 1, 2, 5 and 19 is associated with an increase of structures with bisecting GlcNAc compared to structures without bisecting GlcNAc.

Mediation Analysis

Since the methylation at all 7 CpG-sites associating significantly with IgG glycosylation traits (Figure 1) have been reported to be associated with smoking status [21, 62-67], we tested whether 1) IgG glycosylation is a mediator in the association between smoking and DNA methylation, or whether 2) DNA methylation is a mediator in the association between smoking and IgG glycosylation. For the UPLC data, 6 CpG-sites have been tested for their association with 20 IgG glycan traits (74 associations) in the EGCUT study, while for LC/MS data 3 CpG-sites have been tested for their associations with 2 IgG glycan traits (3 unique associations) in the KORA F4 study.

For the UPLC data in EGCUT, we analyzed 74 IgG glycan – CpG combinations in EGCUT. In total, 64 non-smokers and 84 ever smokers were available. The mediation between smoking and IgG via the CpG-sites (*p*-values range between 2.3×10^{-4} to 0.15, table T2a) is stronger than the mediation between smoking and CpG via IgG (*p*-values range between 8.3×10^{-3} to 0.27, table T2b). The

mediation of the associations between smoking and IGP39 and IGP40 via the CpG-site cg21566642 is Bonferroni-corrected significant ($p < 6.8 \times 10^{-4}$, table T2a).

The mediation analysis in KORA F4 was conducted on 1,594 samples including 671 non-smokers and 932 ever smokers. For all three CpG-sites (cg05951221, cg21566642 and cg06126421), the mediation effect of DNA methylation on the association between smoking and IgG glycosylation was significant (p-values range between 1.3×10^{-8} to 1.4×10^{-3}) (table T2a) while the significance of the mediation effect of the glycans was comparably lower (table T2b). An additional analysis of variance confirmed that the adjustment for the CpG-site in the smoking - IgG glycan model is significant for all 3 CpG-sites ($p = 1.5 \times 10^{-3}$, $p = 1.8 \times 10^{-4}$, $p = 1.3 \times 10^{-8}$). The CpG-sites increase the explained variance of the variation of IgG glycan by 0.52 %, 0.69 % and 0.8 %, respectively (table T2a).

A summarizing plot of the mediation analysis is given in Figure 2. These results hint at a possible mediating effect of DNA methylation on the relation between smoking and IgG glycosylation.

Discussion

In the current paper, we aimed at identifying methylation levels of CpG-sites to be associated with IgG glycosylation traits. We conducted meta-analyses for UPLC IgG glycan data (n = 757) and LC/MS IgG glycan data (n = 2,249) using data from four cohorts (LLS, KORA F4, EGCUT, TwinsUK) and around 3,000 samples contributing to our study. We found 7 CpG-sites associating with IgG glycosylation traits. Out of these, 6 CpG-sites on chromosome 1, 2, 5, and 19 are associated with an increase of structures with bisecting GlcNAc compared to structures without bisecting GlcNAc (UPLC data). Additionally, we found 3 CpG-sites on chromosome 2 and 6 of which a decrease in DNA methylation is associated with an increase of IgG1 fucosylated monosialylated structures with bisecting GlcNAc compared to all fucosylated and monosialylated IgG1 structures (LC/MS data). Two CpG-sites on chromosome 2, cg05951221 and cg21566642, associate with both UPLC as LC/MS IgG glycan traits.

that are suggestive ($p < 1.1 \times 10^{-7}$) in at least one panel (table T1) and show associations with $p < 0.001$ in the other. Out of these, 2 CpG-sites reached the significance threshold of 10^{-7} in both panels. In depth mediation analysis revealed that the effect of smoking on IgG glycosylation levels is mediated by the identified 7 CpG-sites on chromosome 1, 2, 5, 6, and 19.

In fact, all 7 CpG-sites on chromosome 1 (*GNG12*), chromosome 2 (*ALPPL2*), chromosome 5 (*AHRR*), chromosome 6 (6p21.33) and chromosome 19 (*F2RL3*) have been determined to be highly influenced by smoking [21, 62-67]. Strikingly, one of the studies [63] found the genes implicated in the

association of DNA methylation and smoking to be enriched for biological functions related to the immune response, e.g. 'regulation of T-helper 2 cell differentiation'. Similarly, smoking has been shown to reshape the glycemic profile by different analytical methods [68, 69], in particularly including associations to the *N*-glycome [70] and IgG glycosylation [29].

Additionally, smoking has been associated with increased branching of glycan structures [29]. Through the mediation analysis for all 7 CpG-sites, we can hypothesize that smoking leads to a higher ratio between IgG glycan structures with bisecting GlcNAs compared to structures without bisecting GlcNAs, i.e. an increased branching of IgG glycopeptides, via a decrease of DNA methylation at these CpG-sites. For the LC/MS data, we can specify the structure to be additionally fucosylated and monosialylated. None of the loci has been associated to IgG glycosylation before, especially not in a hypothesis-free genome-wide analysis. The effect of these CpG-sites thus may have a specific role in the interaction of smoking and IgG glycosylation [11].

On chromosome 1, the CpG-site cg25189904 is located within the region of *GNG12-AS1*, a GNG12 antisense RNA 1. Its antisense orientation to the gene *DIRAS3* causes the transcriptional silencing of *GNG12-AS1* to upregulate the expression of the tumor-suppressor gene *DIRAS3*, a process which in turn undermines cell cycle progression [71]. However, other studies that found this CpG-site to be highly associated with smoking [21, 62, 63] claim the annotation of cg25189904 to the gene *GNG12* (G protein subunit gamma 12). In addition, the CpG-site is directly annotated to *GNG12* by the manufacturer. Here, further functional studies are recommended to assess the effect of the CpG-site on gene expression and, in a broader sense, on IgG glycosylation.

The 3 CpG-sites on chromosome 2, cg05951221, cg21566642, cg01940273, are located in proximity (233,284,402 – 233,284,402, 2q37.1) of the genes that belong to the group of alkaline phosphatases [61], with *ALPPL2* (chr2:233,271,552-233,275,424) being the closest gene. The encoded glycosylated enzyme is extensively expressed in tumors [72] and could be associated with pancreatic carcinoma [73] and, in one case, with ovarian cancer [73, 74].

Effects for CpG-sites on *AHRR* on chromosome 5, an aryl-hydrocarbon receptor repressor, have been markedly shown in white blood cells [75, 76], including precursors of B-cells which produce immunoglobulin G. Additionally, it could be proven that DNA methylation of cg05575921 in cord blood of newborns is affected by prenatal influences, including maternal smoking behavior and BMI [77, 78]. IgG is transported actively to the fetus via the placenta [79] in order to provide immunological protection for the fetus and newborn. Differences between maternal and cord blood glycosylation for both IgG and the total plasma *N*-glycome have been shown recently [80] and specific *N*-glycans could be associated to a poor fetal environment [81]. A direct link between

glycosylation and DNA methylation in cord blood is however still missing. Given our results, there might be an interesting association.

The gene *F2RL3* on chromosome 19 encodes a protease-activated receptor that plays an important role in blood coagulation and the pathogenesis of inflammation [82] and pain [83]. Its pro-inflammatory effects, however, are still not fully understood [83]. Interestingly, anti- and pro-inflammatory functioning of IgG depends upon its glycosylation structure [1, 4, 5].

For CpG-sites on chromosome 5 (cg05575921, *AHRR*) and on chromosome 19 (cg03636183, *F2RL3*), there is strong evidence that DNA methylation mediates the effects of smoking for cardiovascular diseases [65, 84, 85], all-cause mortality [65, 86, 87] and cancer [88], in particular, for lung cancer [67, 75, 89, 90]. Nevertheless, it has been shown that several challenges including measurement errors, may interfere with the results of mediation analyses [91]. Thus, the results should be handled with caution.

Limitations of our study include the low sample sizes for the mediation analyses, especially for UPLC-measured glycan traits. Additionally, the differing analytical methods for the measurement of IgG glycosylation complicate the joint interpretation of the associated IgG glycan traits. Both analytical methods employed in our analysis, LC/MS and UPLC, bear advantages and disadvantages [2, 46].

Based on our results, we hypothesize that there is a triangular relation between smoking, DNA methylation and IgG glycans, in which DNA methylation mediates the effect of smoking on IgG glycosylation. We detect an increase in IgG glycan branching, especially a relative increase of IgG glycan structures with bisecting GlcNAs that is associated with smoking via downregulated DNA methylation. While it is to a large extent known how smoking affects DNA methylation, the connection to IgG glycans represents a new part of the puzzle explaining the impact of smoking on the human body. Nevertheless our analysis proposes hypotheses here which is a strength of cross-sectional studies. However we are totally aware of primary limitation of the cross-sectional studies, as there is no evidence of temporal relationship between exposure and outcome and has been assessed simultaneously. Without longitudinal data we are not establishing true cause and effect relationship.

In summary, our epigenome-wide analysis of DNA methylation and IgG glycosylation revealed stronger results for UPLC measured data that comprise both Fab and Fc glycosylation sites than for LC/MS measured data including only glycoforms on the Fc part of IgG. Still, all of the associated CpG-sites have been exemplified to be highly influenced by smoking. The conducted mediation analysis

confirms that, at least for two CpG-sites, the effect of smoking on IgG glycosylation, is mediated by DNA methylation. We therefore recommend further studies in order to unfold the interaction of the three components.

List of Tables

List of Figures

Funding

Glycan analysis was supported in part by European Community's Seventh Framework Programme HighGlycan (contract #278535), MIMOmics (contract #305280) and HTP-GlycoMet (contract #324400) grants, as well as the funding for the Croatian National Centre of Research Excellence in Personalised Healthcare.

Acknowledgements

The KORA study was initiated and financed by the Helmholtz Zentrum München – German Research Center for Environmental Health, which is funded by the German Federal Ministry of Education and Research (BMBF) and by the State of Bavaria. Furthermore, KORA research was supported within the Munich Center of Health Sciences (MC-Health), Ludwig-Maximilians-Universität, as part of LMUinnovativ. This study was supported by the German Center for Diabetes Research (DZD e.V.).

The Leiden Longevity Study has received funding from the European Union's Seventh Framework Programme (FP7/2007-2011 and FP7-Health-F5-2012) under grant agreement no. 259679 (IDEAL) and no. 305280 (MIMOmics). This study was supported by a grant from the Innovation-Oriented Research Program on Genomics (SenterNovem IGE05007), the Centre for Medical Systems Biology, and the Netherlands Consortium for Healthy Ageing (grant 050-060-810), all in the framework of the

Netherlands Genomics Initiative, Netherlands Organization for Scientific Research (NWO), Unilever Colworth and by BBMRI-NL, a Research Infrastructure financed by the Dutch government (NWO 184.021.007). J Deelen is financially supported by the Alexander von Humboldt Foundation.

The TwinsUK study was funded by the Wellcome Trust, European Research Council (250157), and European Community's Seventh Framework Programme (FP7/2007-2013). The study also received support from National Institute for Health Research (NIHR) comprehensive Biomedical Research Centre award to Guy's & St Thomas' NHS Foundation Trust in partnership with King's College London. SNP Genotyping was performed by The Wellcome Trust Sanger Institute and National Eye Institute via NIH/CIDR.

The EGCUT study was funded by the Estonian Research Council Grant IUT20-60, EU H2020 grant 692145, and European Union through the European Regional Development Fund (Project No. 2014-2020.4.01.15-0012) GENTRANSMED. We acknowledge EGCUT technical personnel, especially Mr V. Soo and S. Smit. Data analyzes were carried out in part in the High Performance Computing Center of University of Tartu.

Conflict of Interest

G. Lauc is the founder and owner of Genos Ltd, a private research organization that specializes in high-throughput glycomic analysis and has several patents in this field. L. Klarić, J. Štambuk, I. Trbojević-Akmačić, G. Razdorov, F. Vučković, I. Gudelj and J. Krištić are employees of Genos Ltd.

References

1. Arnold, J.N., et al., *The impact of glycosylation on the biological function and structure of human immunoglobulins.* Annu Rev Immunol, 2007. **25**: p. 21-50.
2. Wuhrer, M., et al., *Glycosylation profiling of immunoglobulin G (IgG) subclasses from human serum.* Proteomics, 2007. **7**(22): p. 4070-81.
3. Hamilton, R.G., *The Human IgG Subclasses* ed. C. Mohan. 2001: Calbiochem-Novabiochem Corporation.
4. Kaneko, Y., F. Nimmerjahn, and J.V. Ravetch, *Anti-inflammatory activity of immunoglobulin G resulting from Fc sialylation.* Science, 2006. **313**(5787): p. 670-3.
5. Kobata, A., *The N-linked sugar chains of human immunoglobulin G: their unique pattern, and their functional roles.* Biochim Biophys Acta, 2008. **1780**(3): p. 472-8.
6. van de Bovenkamp, F.S., et al., *The Emerging Importance of IgG Fab Glycosylation in Immunity.* J Immunol, 2016. **196**(4): p. 1435-41.
7. Hang, I., et al., *Analysis of site-specific N-glycan remodelling in the ER and the Golgi.* Glycobiology, 2015. **25**(12): p. 1335–1349.
8. Zoldos, V., et al., *Genomics and epigenomics of the human glycome.* Glycoconj J, 2013. **30**(1): p. 41-50.
9. Pucic, M., et al., *High throughput isolation and glycosylation analysis of IgG-variability and heritability of the IgG glycome in three isolated human populations.* Mol Cell Proteomics, 2011. **10**(10): p. M111 010090.
10. Lauc, G., et al., *Genomics meets glycomics-the first GWAS study of human N-Glycome identifies HNF1alpha as a master regulator of plasma protein fucosylation.* PLoS Genet, 2010. **6**(12): p. e1001256.
11. Lauc, G., et al., *Loci associated with N-glycosylation of human immunoglobulin G show pleiotropy with autoimmune diseases and haematological cancers.* PLoS Genet, 2013. **9**(1): p. e1003225.
12. Menni, C., et al., *Glycosylation of immunoglobulin g: role of genetic and epigenetic influences.* PLoS One, 2013. **8**(12): p. e82558.
13. Lauc, G., J. Kristic, and V. Zoldos, *Glycans - the third revolution in evolution.* Front Genet, 2014. **5**: p. 145.
14. Zoldos, V., S. Grgurevic, and G. Lauc, *Epigenetic regulation of protein glycosylation.* Biomol Concepts, 2010. **1**(3-4): p. 253-61.
15. Lauc, G., A. Vojta, and V. Zoldos, *Epigenetic regulation of glycosylation is the quantum mechanics of biology.* Biochim Biophys Acta, 2014. **1840**(1): p. 65-70.

16. Lee, C.J., et al., *Determining the effect of DNA methylation on gene expression in cancer cells.* Methods Mol Biol, 2014. **1101**: p. 161-78.
17. Suzuki, M.M. and A. Bird, *DNA methylation landscapes: provocative insights from epigenomics.* Nat Rev Genet, 2008. **9**(6): p. 465-76.
18. Bird, A., *Perceptions of epigenetics.* Nature, 2007. **447**(7143): p. 396-8.
19. Han, H., et al., *DNA methylation directly silences genes with non-CpG island promoters and establishes a nucleosome occupied promoter.* Hum Mol Genet, 2011. **20**(22): p. 4299-310.
20. Curradi, M., et al., *Molecular Mechanisms of Gene Silencing Mediated by DNA Methylation.* Molecular and Cellular Biology, 2002. **22**(9): p. 3157-3173.
21. Zeilinger, S., et al., *Tobacco smoking leads to extensive genome-wide changes in DNA methylation.* PLoS One, 2013. **8**(5): p. e63812.
22. Pfeiffer, L., et al., *DNA methylation of lipid-related genes affects blood lipid levels.* Circ Cardiovasc Genet, 2015. **8**(2): p. 334-42.
23. Dekkers, K.F., et al., *Blood lipids influence DNA methylation in circulating cells.* Genome Biol, 2016. **17**(1): p. 138.
24. Voisin, S., et al., *Exercise training and DNA methylation in humans.* Acta Physiol (Oxf), 2015. **213**(1): p. 39-59.
25. Barrios, C., et al., *Glycosylation Profile of IgG in Moderate Kidney Dysfunction.* J Am Soc Nephrol, 2016. **27**(3): p. 933-41.
26. Pinho, S.S. and C.A. Reis, *Glycosylation in cancer: mechanisms and clinical implications.* Nat Rev Cancer, 2015. **15**(9): p. 540-55.
27. Zhang, D., et al., *Disease-specific IgG Fc N-glycosylation as personalized biomarkers to differentiate gastric cancer from benign gastric diseases.* Sci Rep, 2016. **6**: p. 25957.
28. Wang, Y., et al., *The Association Between Glycosylation of Immunoglobulin G and Hypertension: A Multiple Ethnic Cross-Sectional Study.* Medicine (Baltimore), 2016. **95**(17): p. e3379.
29. Knezevic, A., et al., *Effects of aging, body mass index, plasma lipid profiles, and smoking on human plasma N-glycans.* Glycobiology, 2010. **20**(8): p. 959-69.
30. Klasic, M., et al., *DNA hypomethylation upregulates expression of the MGAT3 gene in HepG2 cells and leads to changes in N-glycosylation of secreted glycoproteins.* Sci Rep, 2016. **6**: p. 24363.
31. Greville, G., et al., *Epigenetic regulation of glycosylation and the impact on chemo-resistance in breast and ovarian cancer.* Epigenetics, 2016. **11**(12): p. 845-857.
32. Kizuka, Y. and N. Taniguchi, *Enzymes for N-Glycan Branching and Their Genetic and Nongenetic Regulation in Cancer.* Biomolecules, 2016. **6**(2).

33. Moayyeri, A., et al., *Cohort Profile: TwinsUK and healthy ageing twin study.* Int J Epidemiol, 2013. **42**(1): p. 76-85.
34. Leitsalu, L., et al., *Cohort Profile: Estonian Biobank of the Estonian Genome Center, University of Tartu.* Int J Epidemiol, 2015. **44**(4): p. 1137-47.
35. Holle, R., et al., *KORA--a research platform for population based health research.* Gesundheitswesen, 2005. **67 Suppl 1**: p. S19-25.
36. Schoenmaker, M., et al., *Evidence of genetic enrichment for exceptional survival using a family approach: the Leiden Longevity Study.* Eur J Hum Genet, 2006. **14**(1): p. 79-84.
37. Houseman, E.A., et al., *DNA methylation arrays as surrogate measures of cell mixture distribution.* BMC Bioinformatics, 2012. **13**: p. 86.
38. Westendorp, R.G., et al., *Nonagenarian siblings and their offspring display lower risk of mortality and morbidity than sporadic nonagenarians: The Leiden Longevity Study.* J Am Geriatr Soc, 2009. **57**(9): p. 1634-7.
39. Andrew, T., et al., *Are twins and singletons comparable? A study of disease-related and lifestyle characteristics in adult women.* Twin Res, 2001. **4**(6): p. 464-77.
40. Bibikova, M., et al., *High density DNA methylation array with single CpG site resolution.* Genomics, 2011. **98**(4): p. 288-95.
41. Lehne, B., et al., *A coherent approach for analysis of the Illumina HumanMethylation450 BeadChip improves data quality and performance in epigenome-wide association studies.* Genome Biol, 2015. **16**: p. 37.
42. Teschendorff, A.E., et al., *A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data.* Bioinformatics, 2013. **29**(2): p. 189-96.
43. Bonder, M.J., et al., *Disease variants alter transcription factor levels and methylation of their binding sites.* Nat Genet, 2017. **49**(1): p. 131-138.
44. van Iterson, M., et al., *MethylAid: visual and interactive quality control of large Illumina 450k datasets.* Bioinformatics, 2014. **30**(23): p. 3435-7.
45. Huffman, J.E., et al., *Comparative performance of four methods for high-throughput glycosylation analysis of immunoglobulin G in genetic and epidemiological research.* Mol Cell Proteomics, 2014. **13**(6): p. 1598-610.
46. Selman, M.H., et al., *Fc specific IgG glycosylation profiling by robust nano-reverse phase HPLC-MS using a sheath-flow ESI sprayer interface.* J Proteomics, 2012. **75**(4): p. 1318-29.
47. Keser, T., et al., *Effects of statins on the immunoglobulin G glycome.* Biochim Biophys Acta, 2017. **1861**(5 Pt A): p. 1152-1158.

48. Balbin, M., et al., *DNA sequences specific for Caucasian G3m(b) and (g) allotypes: allotyping at the genomic level.* Immunogenetics, 1994. **39**(3): p. 187-93.
49. Ugrina, I., et al., *glycanr: Tools for Analysing N-Glycan Data (R package version 0.3.0.).* 2016.
50. Johnson, W.E., C. Li, and A. Rabinovic, *Adjusting batch effects in microarray expression data using empirical Bayes methods.* Biostatistics, 2007. **8**(1): p. 118-27.
51. Leek, J.T., et al., *The sva package for removing batch effects and other unwanted variation in high-throughput experiments.* Bioinformatics, 2012. **28**(6): p. 882-3.
52. Beasley, T.M., S. Erickson, and D.B. Allison, *Rank-based inverse normal transformations are increasingly used, but are they merited?* Behav Genet, 2009. **39**(5): p. 580-95.
53. Fabregat-Traver, D., et al., *High-Performance Mixed Models Based Genome-Wide Association Analysis with omicABEL software.* F1000Res, 2014. **3**: p. 200.
54. Wang, J., *cate - High Dimensional Factor Analysis and Confounder Adjusted Testing and Estimation*. 2015.
55. Team, R.C., *R: A Language and Environment for Statistical Computing.* 2015.
56. Viechtbauer, W., *Conducting Meta-Analyses in R with the metafor Package.* Journal of Statistical Software, 2010. **36**(3): p. 1-48.
57. Bonferroni, C.E., *Il calcolo delle assicurazioni su gruppi di teste.* Studi in Onore del Professore Salvatore Ortu Carboni, 1935: p. 13–60.
58. Sobel, M.E., *Asymptotic Confidence Intervals for Indirect Effects in Structural Equation Models.* Sociological Methodology, 1982. **13**: p. 290-312.
59. Preacher, K.J. and A.F. Hayes, *Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models.* Behavior Research Methods, 2008. **40**(3): p. 879-891.
60. Wang, B., *bda: Density Estimation for Grouped Data. R package version 5.1.6.*
61. Martin, D., et al., *Comparison of the three PLAP-related genes on human chromosome 2.* Clin Chim Acta, 1990. **186**(2): p. 165-70.
62. Tsaprouni, L.G., et al., *Cigarette smoking reduces DNA methylation levels at multiple genomic loci but the effect is partially reversible upon cessation.* Epigenetics, 2014. **9**(10): p. 1382-96.
63. Besingi, W. and A. Johansson, *Smoke-related DNA methylation changes in the etiology of human disease.* Hum Mol Genet, 2014. **23**(9): p. 2290-7.
64. Bauer, M., et al., *Tobacco smoking differently influences cell types of the innate and adaptive immune system-indications from CpG site methylation.* Clin Epigenetics, 2015. **7**: p. 83.
65. Zhang, Y., et al., *Smoking-Associated DNA Methylation Biomarkers and Their Predictive Value for All-Cause and Cardiovascular Mortality.* Environ Health Perspect, 2016. **124**(1): p. 67-74.

66. Shenker, N.S., et al., *Epigenome-wide association study in the European Prospective Investigation into Cancer and Nutrition (EPIC-Turin) identifies novel genetic loci associated with smoking.* Hum Mol Genet, 2013. **22**(5): p. 843-51.

67. Baglietto, L., et al., *DNA methylation changes measured in pre-diagnostic peripheral blood samples are associated with smoking and lung cancer risk.* Int J Cancer, 2017. **140**(1): p. 50-61.

68. Arnold, J.N., et al., *Novel glycan biomarkers for the detection of lung cancer.* J Proteome Res, 2011. **10**(4): p. 1755-64.

69. Vasseur, J.A., et al., *Smoking and lung cancer-induced changes in N-glycosylation of blood serum proteins.* Glycobiology, 2012. **22**(12): p. 1684-708.

70. Reiding, K.R., et al., *Human Plasma N-glycosylation as Analyzed by Matrix-Assisted Laser Desorption/Ionization-Fourier Transform Ion Cyclotron Resonance-MS Associates with Markers of Inflammation and Metabolic Health.* Mol Cell Proteomics, 2017. **16**(2): p. 228-242.

71. Stojic, L., et al., *Transcriptional silencing of long noncoding RNA GNG12-AS1 uncouples its transcriptional and product-related functions.* Nat Commun, 2016. **7**: p. 10406.

72. Le Du, M.H., et al., *Crystal structure of alkaline phosphatase from human placenta at 1.8 A resolution. Implication for a substrate specificity.* J Biol Chem, 2001. **276**(12): p. 9158-65.

73. Dua, P., et al., *Alkaline phosphatase ALPPL-2 is a novel pancreatic carcinoma-associated protein.* Cancer Res, 2013. **73**(6): p. 1934-45.

74. Ben-Arie, A., et al., *Elevated serum alkaline phosphatase may enable early diagnosis of ovarian cancer.* Eur J Obstet Gynecol Reprod Biol, 1999. **86**(1): p. 69-71.

75. Monick, M.M., et al., *Coordinated changes in AHRR methylation in lymphoblasts and pulmonary macrophages from smokers.* Am J Med Genet B Neuropsychiatr Genet, 2012. **159B**(2): p. 141-51.

76. Philibert, R.A., S.R. Beach, and G.H. Brody, *Demethylation of the aryl hydrocarbon receptor repressor as a biomarker for nascent smokers.* Epigenetics, 2012. **7**(11): p. 1331-8.

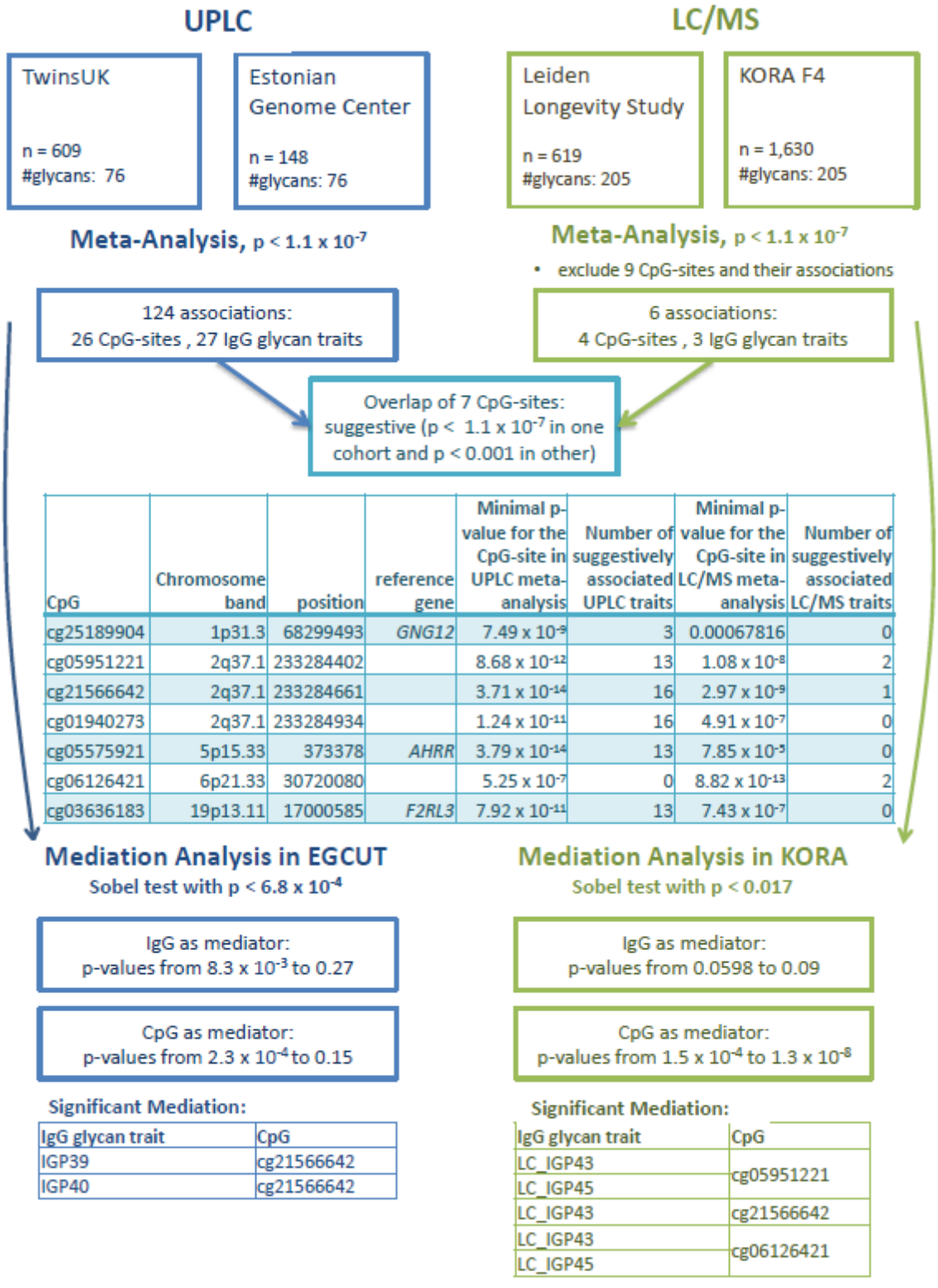
77. Kupers, L.K., et al., *DNA methylation mediates the effect of maternal smoking during pregnancy on birthweight of the offspring.* Int J Epidemiol, 2015. **44**(4): p. 1224-37.

78. Burris, H.H., et al., *Offspring DNA methylation of the aryl-hydrocarbon receptor repressor gene is associated with maternal BMI, gestational age, and birth weight.* Epigenetics, 2015. **10**(10): p. 913-21.

79. Simister, N.E., et al., *An IgG-transporting Fc receptor expressed in the syncytiotrophoblast of human placenta.* Eur J Immunol, 1996. **26**(7): p. 1527-31.

80. Jansen, B.C., et al., *MALDI-TOF-MS reveals differential N-linked plasma- and IgG-glycosylation profiles between mothers and their newborns.* Sci Rep, 2016. **6**: p. 34001.
81. Sato, R., et al., *Fetal Environment and Glycosylation Status in Neonatal Cord Blood.* Medicine, 2016. **95**(14): p. e3219.
82. Fu, Q., et al., *Protease-activated receptor 4: a critical participator in inflammatory response.* Inflammation, 2015. **38**(2): p. 886-95.
83. Bao, Y., et al., *New insights into protease-activated receptor 4 signaling pathways in the pathogenesis of inflammation and neuropathic pain: a literature review.* Channels (Austin), 2015. **9**(1): p. 5-13.
84. Breitling, L.P., et al., *Smoking, F2RL3 methylation, and prognosis in stable coronary heart disease.* Eur Heart J, 2012. **33**(22): p. 2841-8.
85. Jhun, M.A., *Epidemiologic approaches to understanding mechanisms of cardiovascular diseases: Genes, environment, and DNA methylation (Doctoral Thesis).* 2015.
86. Bojesen, S.E., et al., *AHRR (cg05575921) hypomethylation marks smoking behaviour, morbidity and mortality.* Thorax, 2017.
87. Zhang, Y., et al., *F2RL3 methylation in blood DNA is a strong predictor of mortality.* Int J Epidemiol, 2014. **43**(4): p. 1215-25.
88. Yu, G., et al., *Increased expression of protease-activated receptor 4 and Trefoil factor 2 in human colorectal cancer.* PLoS One, 2015. **10**(4): p. e0122678.
89. Fasanelli, F., et al., *Hypomethylation of smoking-related genes is associated with future lung cancer in four prospective cohorts.* Nat Commun, 2015. **6**: p. 10192.
90. Zhang, Y., et al., *F2RL3 methylation, lung cancer incidence and mortality.* Int J Cancer, 2015. **137**(7): p. 1739-48.
91. Richmond, R.C., et al., *Challenges and novel approaches for investigating molecular mediation.* Hum Mol Genet, 2016. **25**(R2): p. R149-R156.

Figure 1



CpG	Chromosome band	position	reference gene	Minimal p-value for the CpG-site in UPLC meta-analysis	Number of suggestively associated UPLC traits	Minimal p-value for the CpG-site in LC/MS meta-analysis	Number of suggestively associated LC/MS traits
cg25189904	1p31.3	68299493	*GNG12*	7.49×10^{-9}	3	0.00067816	0
cg05951221	2q37.1	233284402		8.68×10^{-12}	13	1.08×10^{-8}	2
cg21566642	2q37.1	233284661		3.71×10^{-14}	16	2.97×10^{-9}	1
cg01940273	2q37.1	233284934		1.24×10^{-11}	16	4.91×10^{-7}	0
cg05575921	5p15.33	373378	*AHRR*	3.79×10^{-14}	13	7.85×10^{-5}	0
cg06126421	6p21.33	30720080		5.25×10^{-7}	0	8.82×10^{-13}	2
cg03636183	19p13.11	17000585	*F2RL3*	7.92×10^{-11}	13	7.43×10^{-7}	0

Figure 2

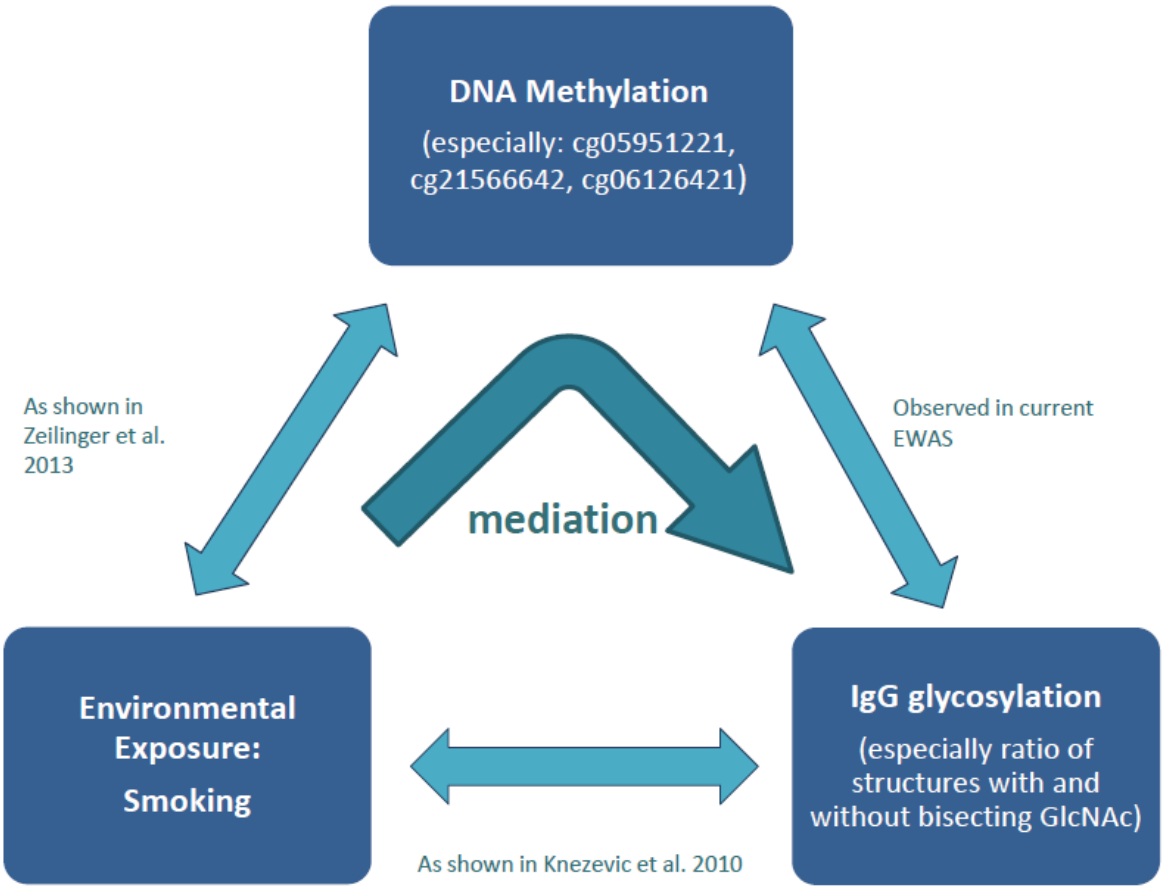


Table 1

Overview on suggestive CpG-sites form one panel and p < 0.001 in the other

CpG	Chromosome	position	reference gene	Minimal p-value for the CpG-site in UPLC meta-analysis	Number of suggestively associated UPLC traits	Associated UPLC IgG Glycan Traits	Minimal p-value for the CpG-site in LC/MS meta-analysis	Number of suggestively associated LC/MS traits	Associated LC/MS IgG Glycan Traits
cg25189904	1p31.3	68299493	GNG12	7.49E-09	3	IGP49, GP10, IGP68	6.78E-04	0	
cg05951221	2q37.1	233284402		8.68E-12	13	IGP66, IGP68, IGP62, GP10, IGP63, IGP70, IGP72, IGP67, IGP39, IGP40, IGP49, IGP71, IGP64	1.08E-08	2	LC_IGP43, LC_IGP45
cg21566642	2q37.1	233284661		3.71E-14	16	IGP66, IGP75, IGP74, IGP69, IGP62, IGP71, IGP65, IGP39, GP10, IGP76, IGP70, IGP40, IGP49, IGP68, IGP64, IGP72	2.97E-09	1	LC_IGP43
cg01940273	2q37.1	233284934		1.24E-11	16	IGP70, IGP39, IGP62, IGP68, IGP40, IGP67, IGP75, IGP64, IGP66, IGP63, IGP71, IGP74, IGP69, GP10, IGP49, IGP72	4.91E-07	0	

cg0557591	5p15.33	3733778	AHRR	3.79E-14	13	GP23, GP10, IGP64, IGP70, IGP34, IGP39, IGP40, IGP62, IGP66, IGP68, IGP71, IGP72, IGP49	7.85E-05	0	
cg0612641	6p21.33	3072080		5.25E-07	0		8.82E-13	2	LC_IGP43, LC_IGP45
cg0363618	19p13.11	1700585	F2RL3	7.92E-11	13	IGP62, IGP63, IGP72, IGP71, IGP67, IGP40, IGP39, IGP49, IGP66, IGP64, IGP70, GP10, IGP68	7.43E-07	0	

Table 2a

Results for Mediation Analysis in KORA and EGCUT with the CpG-site as potential mediator

CpG as Mediator

LC/MS: threshold = 0.017			Anova: IgG ~ CpG + smoking + covariates	
IgG glycan trait	CpG	p (Sobel-test)	Additional Variance Explained by the CpG-site (R2)	p-value of CpG-site
LC_IGP43	cg05951221	0.001485682	0.005210657	0.002618943
LC_IGP45	cg05951221	0.001485682	0.005210657	0.002618943
LC_IGP43	cg21566642	0.000177866	0.006896529	0.000532924
LC_IGP43	cg06126421	1.31713E-08	0.008021274	0.00018653
LC_IGP45	cg06126421	1.31713E-08	0.008021274	0.00018653

UPLC: threshold = 6.8e-04			Anova: IgG ~ CpG + smoking + covariates	
IgG glycan trait	CpG	p (Sobel-test)	Additional Variance Explained by the CpG-site (R2)	p-value of CpG-site
IGP49	cg25189904	0.062994361	0.05416807	0.002419576
GP10	cg25189904	0.061991431	0.058431596	0.00158616
IGP68	cg25189904	0.014448871	0.072188424	0.000291678
IGP66	cg05951221	0.022542464	0.031150223	0.020818396
IGP68	cg05951221	0.004971857	0.041564239	0.006627828
IGP62	cg05951221	0.033277717	0.036814666	0.014137276
GP10	cg05951221	0.044336053	0.022948982	0.050797773
IGP63	cg05951221	0.098521911	0.026457082	0.045463986
IGP70	cg05951221	0.026530856	0.030814716	0.021918835
IGP72	cg05951221	0.025005668	0.030895214	0.021911266
IGP67	cg05951221	0.125752028	0.015437522	0.119278616
IGP39	cg05951221	0.009511351	0.026275257	0.021771215
IGP40	cg05951221	0.009511351	0.026275257	0.021771215
IGP49	cg05951221	0.035970392	0.022213343	0.054940179
IGP71	cg05951221	0.026530856	0.030814716	0.021918835

IGP64	cg05951221	0.005114167	0.049276526	0.003428988
IGP66	cg21566642	0.006203155	0.035856714	0.012984802
IGP75	cg21566642	0.005538572	0.037931321	0.010702606
IGP74	cg21566642	0.005538572	0.037931321	0.010702606
IGP69	cg21566642	0.00722555	0.035384995	0.014334399
IGP62	cg21566642	0.008943166	0.044057296	0.007115189
IGP71	cg21566642	0.007351238	0.035691951	0.013471837
IGP65	cg21566642	0.006878152	0.053019324	0.002910797
IGP39	cg21566642	0.000236086	0.038360744	0.005315167
GP10	cg21566642	0.03613135	0.024624397	0.042921596
IGP76	cg21566642	0.009187717	0.035152957	0.014955653
IGP70	cg21566642	0.007351238	0.035691951	0.013471837
IGP40	cg21566642	0.000236086	0.038360744	0.005315167
IGP49	cg21566642	0.028970152	0.02431183	0.044483128
IGP68	cg21566642	0.001709601	0.043391679	0.00550254
IGP64	cg21566642	0.001819179	0.050451186	0.00305211
IGP72	cg21566642	0.007768163	0.034891509	0.014715385
IGP70	cg01940273	0.013437126	0.030102714	0.02354151
IGP39	cg01940273	0.010470556	0.021393124	0.038850408
IGP62	cg01940273	0.008795505	0.037359917	0.013422772
IGP68	cg01940273	0.005353127	0.035522961	0.012281103
IGP40	cg01940273	0.010470556	0.021393124	0.038850408
IGP67	cg01940273	0.062134358	0.018057553	0.091727474
IGP75	cg01940273	0.041690728	0.015360187	0.107314591
IGP64	cg01940273	0.004359403	0.039945194	0.008657255
IGP66	cg01940273	0.01323863	0.030761408	0.021650035
IGP63	cg01940273	0.025732389	0.03198015	0.027572525
IGP71	cg01940273	0.013437126	0.030102714	0.02354151
IGP74	cg01940273	0.041690728	0.015360187	0.107314591
IGP69	cg01940273	0.064934023	0.013377681	0.135331853
GP10	cg01940273	0.034666225	0.023150314	0.049776663
IGP49	cg01940273	0.025878659	0.023509451	0.048212699
IGP72	cg01940273	0.013928045	0.029413125	0.025417003
GP23	cg05575921	0.151679974	0.03157825	0.019449147

GP10	cg05575921	0.003856893	0.049063578	0.00392826
IGP64	cg05575921	0.001173097	0.055445413	0.001860242
IGP70	cg05575921	0.004529522	0.03891391	0.009784855
IGP34	cg05575921	0.004731005	0.046174244	0.004448482
IGP39	cg05575921	0.024012772	0.016151229	0.073325842
IGP40	cg05575921	0.024012772	0.016151229	0.073325842
IGP62	cg05575921	0.010054398	0.041846723	0.008770027
IGP66	cg05575921	0.003995263	0.038752123	0.009726404
IGP68	cg05575921	0.000804767	0.052738094	0.002127067
IGP71	cg05575921	0.004529522	0.03891391	0.009784855
IGP72	cg05575921	0.004621036	0.038931538	0.009862206
IGP49	cg05575921	0.002434214	0.045591847	0.005541803
IGP62	cg03636183	0.056134614	0.014525275	0.126262611
IGP63	cg03636183	0.050343218	0.017509706	0.104701109
IGP72	cg03636183	0.029463857	0.01801353	0.081572755
IGP71	cg03636183	0.029586753	0.018398182	0.077966265
IGP67	cg03636183	0.024531846	0.020042005	0.075432295
IGP40	cg03636183	0.139397648	0.013089355	0.107401217
IGP39	cg03636183	0.139397648	0.013089355	0.107401217
IGP49	cg03636183	0.007802105	0.035516613	0.014731495
IGP66	cg03636183	0.026773576	0.019837948	0.066271513
IGP64	cg03636183	0.016032079	0.02476639	0.039853874
IGP70	cg03636183	0.029586753	0.018398182	0.077966265
GP10	cg03636183	0.01475222	0.035095594	0.015254734
IGP68	cg03636183	0.010868436	0.029333674	0.023214205

Table 2b

Results for Mediation Analysis in KORA and EGCUT with the IgG glycan as potential mediator

IgG as Mediator

LC/MS: threshold = 0.017			Anova: CpG ~ IgG + smoking + covariates	
IgG glycan trait	CpG	p (Sobel-test)	Additional Variance Explained by the IgG glycan traits (R2)	p-value of the IgG glycan trait
LC_IGP43	cg05951221	0.091766616	0.004234085	0.002618943
LC_IGP45	cg05951221	0.091766616	0.004234085	0.002618943
LC_IGP43	cg21566642	0.078966329	0.005259427	0.000532924
LC_IGP43	cg06126421	0.059804893	0.005910232	0.00018653
LC_IGP45	cg06126421	0.059804893	0.005910232	0.00018653

UPLC: threshold = 6.8e-04			Anova: CpG ~ IgG + smoking + covariates	
IgG glycan trait	CpG	p (Sobel-test)	Additional Variance Explained by the IgG glycan traits (R2)	p-value of the IgG glycan trait
IGP49	cg25189904	0.143197057	0.047336711	0.002419576
GP10	cg25189904	0.158870739	0.051154534	0.00158616
IGP68	cg25189904	0.035416491	0.066390015	0.000291678
IGP66	cg05951221	0.063062909	0.016891058	0.020818396
IGP68	cg05951221	0.024648278	0.023103555	0.006627828
IGP62	cg05951221	0.083614538	0.018982066	0.014137276
GP10	cg05951221	0.14743234	0.012142933	0.050797773
IGP63	cg05951221	0.20876835	0.012725983	0.045463986
IGP70	cg05951221	0.067909783	0.016613844	0.021918835
IGP72	cg05951221	0.067515719	0.016615701	0.021911266
IGP67	cg05951221	0.208154303	0.007771197	0.119278616
IGP39	cg05951221	0.032333138	0.016650195	0.021771215
IGP40	cg05951221	0.032333138	0.016650195	0.021771215

IGP49	cg05951221	0.12154544	0.01173255	0.054940179
IGP71	cg05951221	0.067909783	0.016613844	0.021918835
IGP64	cg05951221	0.025891032	0.026703995	0.003428988
IGP66	cg21566642	0.037347592	0.01644872	0.012984802
IGP75	cg21566642	0.022716844	0.017336831	0.010702606
IGP74	cg21566642	0.022716844	0.017336831	0.010702606
IGP69	cg21566642	0.026375984	0.01599519	0.014334399
IGP62	cg21566642	0.049717893	0.019217921	0.007115189
IGP71	cg21566642	0.039527675	0.016279767	0.013471837
IGP65	cg21566642	0.044260365	0.023349157	0.002910797
IGP39	cg21566642	0.008294586	0.020564839	0.005315167
GP10	cg21566642	0.139027246	0.011022794	0.042921596
IGP76	cg21566642	0.038090025	0.01580078	0.014955653
IGP70	cg21566642	0.039527675	0.016279767	0.013471837
IGP40	cg21566642	0.008294586	0.020564839	0.005315167
IGP49	cg21566642	0.113575514	0.010863308	0.044483128
IGP68	cg21566642	0.016183057	0.020404755	0.00550254
IGP64	cg21566642	0.017360292	0.023129885	0.00305211
IGP72	cg21566642	0.041727695	0.015874979	0.014715385
IGP70	cg01940273	0.050349502	0.015743813	0.02354151
IGP39	cg01940273	0.034132643	0.013150397	0.038850408
IGP62	cg01940273	0.049738823	0.018686194	0.013422772
IGP68	cg01940273	0.025673242	0.019154043	0.012281103
IGP40	cg01940273	0.034132643	0.013150397	0.038850408
IGP67	cg01940273	0.154241798	0.008817826	0.091727474
IGP75	cg01940273	0.069952612	0.008049977	0.107314591
IGP64	cg01940273	0.024461891	0.020998729	0.008657255
IGP66	cg01940273	0.050226093	0.016180585	0.021650035
IGP63	cg01940273	0.143897277	0.014921834	0.027572525
IGP71	cg01940273	0.050349502	0.015743813	0.02354151
IGP74	cg01940273	0.069952612	0.008049977	0.107314591
IGP69	cg01940273	0.094859733	0.006933898	0.135331853
GP10	cg01940273	0.13776097	0.011882542	0.049776663
IGP49	cg01940273	0.110158559	0.01204518	0.048212699

IGP72	cg01940273	0.05256029	0.01534479	0.025417003
GP23	cg05575921	0.269703323	0.017421455	0.019449147
GP10	cg05575921	0.089344319	0.026207319	0.00392826
IGP64	cg05575921	0.014218745	0.030332276	0.001860242
IGP70	cg05575921	0.03238985	0.021179712	0.009784855
IGP34	cg05575921	0.047010261	0.025521109	0.004448482
IGP39	cg05575921	0.051370285	0.010331923	0.073325842
IGP40	cg05575921	0.051370285	0.010331923	0.073325842
IGP62	cg05575921	0.050680629	0.021781476	0.008770027
IGP66	cg05575921	0.031212684	0.021212625	0.009726404
IGP68	cg05575921	0.011848671	0.029592839	0.002127067
IGP71	cg05575921	0.03238985	0.021179712	0.009784855
IGP72	cg05575921	0.033922796	0.02113646	0.009862206
IGP49	cg05575921	0.066113572	0.024309101	0.005541803
IGP62	cg03636183	0.102587704	0.009467955	0.126262611
IGP63	cg03636183	0.163628154	0.01064725	0.104701109
IGP72	cg03636183	0.067341818	0.012247163	0.081572755
IGP71	cg03636183	0.065984785	0.012539957	0.077966265
IGP67	cg03636183	0.109371773	0.012754419	0.075432295
IGP40	cg03636183	0.16350642	0.010485757	0.107401217
IGP39	cg03636183	0.16350642	0.010485757	0.107401217
IGP49	cg03636183	0.076541736	0.023714787	0.014731495
IGP66	cg03636183	0.062792499	0.013598833	0.066271513
IGP64	cg03636183	0.038739084	0.016967117	0.039853874
IGP70	cg03636183	0.065984785	0.012539957	0.077966265
GP10	cg03636183	0.107093929	0.023475884	0.015254734
IGP68	cg03636183	0.030339903	0.020612689	0.023214205