

**Replication of fifteen loci involved in human plasma protein *N*-glycosylation in  
4,802 samples from four cohorts**

**Supplementary Methods**

**Table of contents**

|                                                 |   |
|-------------------------------------------------|---|
| Study cohort description .....                  | 2 |
| EPIC-Potsdam .....                              | 2 |
| PainOmics .....                                 | 2 |
| SOCCS .....                                     | 2 |
| SABRE .....                                     | 3 |
| Genotyping .....                                | 3 |
| EPIC-Potsdam .....                              | 3 |
| PainOmics .....                                 | 3 |
| SOCCS .....                                     | 4 |
| SABRE .....                                     | 4 |
| Phenotyping .....                               | 4 |
| Plasma N-glycome quantification .....           | 4 |
| Harmonization of glycan peaks .....             | 5 |
| Normalization and batch-correction of GPs ..... | 5 |
| Association analysis and meta-analysis .....    | 5 |
| Replication .....                               | 6 |
| Statistical power calculation .....             | 6 |
| References .....                                | 7 |

## **Study cohort description**

### ***EPIC-Potsdam***

The prospective EPIC-Potsdam cohort study includes 27,548 participants (16,644 women and 10,904 men), who were recruited within an age-range of 35-65 years from the general population between 1994 and 1998 (Boeing, Korfmann, et al. 1999). Baseline assessments comprised anthropometric measures and blood sampling by qualified medical personnel. Blood pressure was measured with the individual sitting with the arm elevated at heart level by using oscillometric devices (BOSO- Oscillomat; Bosch & Sohn, Jungingen, Germany), and the mean of the second and third reading was used. Lifestyle, sociodemographic characteristics, and current health status were assessed with validated questionnaires and in face-to-face interviews (Boeing, Wahrendorf, et al. 1999). Participants were then actively contacted by sending out questionnaires and, if necessary, by telephone every 2–3 years with response rates between 90 and 96% per follow-up round (Bergmann et al. 1999). From all participants who provided blood at baseline (n=26,437), a random sample (subcohort, n=2,500) was drawn, which served as study population for the present analysis. All participants gave written informed consent for biomedical research, and the study was approved by the Ethics Committee of the State of Brandenburg, Germany (Boeing, Korfmann, et al. 1999).

### ***PainOmics***

The PainOmics (Allegri et al. 2016) study of patients comprised a retrospective multicenter study (ClinicalTrials.gov Identifier NCT02037789) part of the PainOMICS project funded by European Community in the Seventh Framework Programme (Project ID: 602736). The primary objective is to recognize genetic variants associated with chronic low back pain (CLBP); secondary objectives are to study glycomic and proteomic profiles associated with CLBP. Glycomic and Activomic approaches aim to reveal alterations in proteome complexity that arise from the post-translational modification that varies in response to changes in the physiological environment, a particularly important avenue to explore in chronic inflammatory diseases. The study was firstly approved by the Institutional Review Boards of IRCCS Foundation San Matteo Hospital Pavia and then by the Institutional Review boards of all clinical centers (King's College London, ZOL Genk/Lanaken, St. Catherine Specialty Hospital) that enrolled patients. Copies of approvals were provided to the European Commission before starting the study. Written informed consent was obtained from all participants. In the period between September 2014 and February 2016, 3,400 patients were enrolled. Sample collection was performed in all patients enrolled, according to the Standard Operating Procedures published in PlosOne in 2017 (Dagostino et al. 2017). Samples were processed in the PainOmics laboratory in a blinded manner in the University of Parma. Blood collections for omics analyses were taken at the time of enrollment. All blood samples were processed and stored according to SOP protocols until they were sent to analytical laboratories for omics analysis.

### ***SOCCS***

SOCCS study (Theodoratou et al. 2016; Vuckovic et al. 2016) comprised 2,057 colorectal cancer cases (CRC) (61% male; mean age at diagnosis ( $65.8 \pm 8.4$ ) years) and 2,111 population controls (60% males; mean age  $67.9 \pm 9.0$  years) as ascertained in Scotland. Cases were taken from

an independent prospective incident CRC case series and were aged < 80 years at diagnosis. Control subjects were population controls matched by age ( $\pm 5$  years), gender, and area of residence within Scotland. All participants gave written informed consent and study approval was from the MultiCentre Research Ethics Committee for Scotland and Local Research Ethics committee. Sample collection is described in (Theodoratou et al. 2016; Vuckovic et al. 2016).

### **SABRE**

SABRE is a population-based cohort initiated in 4,857 people of European, South Asian and African Caribbean origin aged 40 to 69 living in West London, UK (Tillin et al. 2012). Peripheral blood samples were collected from the Southall participants at baseline (1988–91) for DNA extraction. We used data from the European-origin participants. Plasma *N*-glycome quantification was performed using samples collected at the 20 year follow-up visit (2008–2011). All participants gave written informed consent. Approval for the baseline study was obtained from Ealing, Hounslow, and Spelthorne, Parkside and University College London research ethnics committees.

### **Genotyping**

Genotyping was carried out using the SNP-array with genome-wide coverage, followed by genomic imputation. The genotype data quality control was conducted using the protocol, guidelines and software, described in (Anderson et al. 2010).

### **EPIC-Potsdam**

Genotyping was carried out using three different genotyping arrays: Human660W-Quad\_v1\_A ( $N = 325$ ), HumanCoreExome-12v1-0\_B ( $N = 622$ ) and Illumina InfiniumOmniExpressExome-8v1-3\_ADNA Analysis BeadChip ( $N = 1,245$ ). Genotyping and quality control of the Human660W-Quad\_v1\_A and HumanCoreExome-12v1-0\_B chips was described elsewhere (Langenberg et al. 2014). Genotype calling and quality control of the samples genotyped with InfiniumOmniExpressExome-8v1-3\_ADNA Analysis BeadChip were carried out jointly using Illuminas GenomeStudio v2011.1 software suite and protocols suggested by the CHARGE consortium (Grove et al. 2013), Anderson et al (Anderson et al. 2010) and Guo et al. (Guo et al. 2014). To improve the genotype calling for rare variants zCall with a threshold of 7 was applied (Goldstein et al. 2012). Individuals with low call rate, discordant sex information ( $F$  value between 0.2 and 0.8), related or duplicated individuals ( $IBD > 0.185$ ) and individuals with divergent ancestry were excluded from further analysis. Phasing and imputation were conducted using the Michigan Imputation Service (Das et al. 2016). The Haplotype Reference Consortium (release 1.1) was used as reference panel (McCarthy et al. 2016). Before imputation pre-phasing was applied using Eagle2 (Loh et al. 2016). Imputation was carried out in four separated datasets (one for each genotyping chip or two for the HumanCoreExome-12v1-0\_B chip) using minimac3 (Das et al. 2016). Genotypes map to the GRCh37 human genome build.

### **PainOmics**

Genotyping was carried out using Illumina HumanCore BeadChip and Illumina GSA. Quality control (QC) of genotype was conducted separately for each SNP-array. Standard QC of genotyped data was applied with SNPs filtered by a sample call rate > 98%, minor allele frequency (MAF) > 0.625%, SNP call rate: > 97%, Hardy-Weinberg equilibrium (HWE)  $P \leq 1 \times 10^{-6}$ .

Imputation was done using the Eagle software tool with HRC r1.1 2016 reference and mapped to the GRCh37 human genome build.

### **SOCCS**

Details of the genotyping procedure can be found in (Tenesa et al. 2008) and in (Law et al. 2019). Genotyping was carried out using Illumina SNP arrays HumanHap300, HumanHap240S or Infinium OmniExpressExome-8. Standard QC of genotype data was applied, with SNPs filtered by sample call rate > 95%, MAF > 1%, SNP call rate in autosomal chromosomes > 99%, HWE  $P \leq$  Bonferonni cut-off > 0.05/number of tested SNPs. In total, 514,177 SNPs passed these criteria. Imputation was done using SHAPEIT and IMPUTE2 software with 1000 Genomes, phase 1 (Integrated haplotypes, released June 2014) and mapped to the GRCh37 human genome build.

### **SABRE**

Genotyping was done using the Illumina Human Core Bead Chip. In total, 1429 samples of European descent were genotyped. QC of genotype data was conducted separately for each ethnic group. Standard QC of genotype data was applied, with SNPs filtered by sample call rate > 95%, MAF > 1%, SNP call rate > 95%, HWE  $P \leq 1 \times 10^{-4}$ . In total, 332,849 SNPs passed these criteria. Imputation was done with Michigan imputation server using Eagle and Minimac software. HRC reference panel was used for imputation of samples with European descent. The GRCh37 human genome build was used.

## **Phenotyping**

### ***Plasma N-glycome quantification***

Plasma N-glycome quantification of samples from the EPIC-Potsdam, PainOmics, and SABRE studies was performed at GENOS by applying the following protocol. Plasma N-glycans were enzymatically released from proteins by PNGase F, fluorescently labeled with 2-aminobenzamide and cleaned-up from the excess of reagents by hydrophilic interaction liquid chromatography solid phase extraction (HILIC-SPE) as described previously (Trbojević Akmačić et al. 2015). Fluorescently labeled and purified N-glycans were separated by HILIC on a Waters BEH Glycan chromatography column, 150 × 2.1 mm, 1.7 µm BEH particles, installed on an Acquity ultra-high-performance liquid chromatography (UHPLC) instrument (Waters, Milford, MA, USA) consisting of a quaternary solvent manager, sample manager and a fluorescence detector set with excitation and emission wavelengths of 250 nm and 428 nm, respectively. Following chromatography conditions previously described in detail (Trbojević Akmačić et al. 2015), glycan peaks (GPs) – quantitative measurements of glycan levels – were defined by manual integration of intensity peaks in the chromatograms for PainOmics retrospective study (Italian subcohort) and for SABRE cohort, while for other PainOmics subcohorts and EPIC-Potsdam GPs were defined by automatic integration (Agakova et al. 2017). The number of defined GPs varied among studies from 36 to 42. The abundance of N-glycans in each chromatographic peak was expressed as percentage area of the corresponding peak.

Plasma N-glycome quantification for SOCCS samples was done at NIBRT by applying the same protocol as for EPIC-Potsdam, PainOmics, and SABRE with the only difference in the excitation wavelength (330 nm instead of 250 nm).

### ***Harmonization of glycan peaks***

Depending on the cohort, some neighboring chromatographic peaks were not well separated due to differences between used UHPLC column batches and were consequently quantified as one chromatographic peak. To conduct association analysis in four cohorts followed by meta-analysis, we harmonized the set of GPs by applying a recently published protocol (Sharapov et al. 2019). In short, based on this table of correspondence between GPs across cohorts (see Table SIV), we created a harmonization scheme. Applying this procedure resulted in a harmonized set of 36 GPs for all samples.

### ***Normalization and batch-correction of GPs***

Normalization and batch-correction were performed on harmonized UHPLC glycan data separately for four cohorts: EPIC-Potsdam, PainOmics, SOCCS, and SABRE. We used probabilistic median quotient normalization (Dieterle et al. 2006; Benedetti et al. 2019 Oct 24). This approach is based on the calculation of the dilution factor of each sample with respect to a reference sample. The reference sample was calculated as the median value of the abundance of each GP across all measured samples. For each sample, a vector of quotients was then obtained by dividing each GP measurement by the corresponding value in the reference sample. The median of these quotients was used as the sample's dilution factor, and the original sample values divided by that value. The underlying assumption is that the different intensities observed across individuals relate to different amounts of the biological material in the collected samples.

Normalized glycan measurements were log10-transformed due to the right skewness of distribution and multiplicative nature of batch effects. Before the batch correction, outlying measurements were removed. An outlier was defined as a sample that had at least one GP that is  $> 3$  standard deviations from the mean. The batch correction was performed on log10-transformed measurements using the ComBat method (Johnson et al. 2007), where the technical source of variation (batch and plate number) was modeled as a batch covariate. Again, samples with outlying measurements were removed.

From the 36 directly measured glycan traits, 81 derived traits were calculated (see Table SI). These derived traits average glycosylation features such as branching, galactosylation, and sialylation, etc across different individual glycan structures and, consequently, they may be more closely related to individual enzymatic activity and underlying genetic polymorphism. As derived traits represent sums of directly measured glycans, they were calculated using normalized and batch-corrected glycan measurements after transformation to the proportions (exponential transformation of batch-corrected measurements).

Before the association analysis, the total plasma *N*-glycome traits were adjusted for sex and age. Residuals were quantile transformed to normal distribution (rnttransform function in the GenABEL v. 1.8-0 (Aulchenko et al. 2007; Karssen et al. 2016) R package).

### ***Association analysis and meta-analysis***

The protocols of association analysis and meta-analysis were created based on the published protocol (Winkler et al. 2014), where protocols, pipelines, code usage, and 'toy' data are provided and extensively described. We performed association analysis for tag SNPs, representing sixteen loci, previously reported to be associated with total plasma *N*-glycome traits (Lauc et al. 2010;

Huffman et al. 2011; Sharapov et al. 2019). Tag SNPs for the loci near the genes *FUT3/FUT5/FUT6*, *FUT8*, *HNF1a*, *B3GAT1*, *MGAT5*, and *SLC9A9*, associations of which raised in the series of GWAS conducted on HPLC-measured total plasma *N*-glycome loci, were selected by choosing the SNP with the strongest association as reported by Sharapov et al. (Sharapov et al. 2019). Tag SNPs for other loci, associations of which raised in the first GWAS of UPLC-measured total plasma *N*-glycome loci, were selected by choosing the SNP with the strongest association as reported by Huffman et al. (Huffman et al. 2011). The rs59111563 SNP tagging *ST6GAL1* locus was not present in our replication set, and we, therefore, replaced it by rs17775791 (LD  $r^2 = 0.9898$ , as calculated by LDLink, EUR subset of samples from 1000 Genomes phase 3 version 5, allele rs59111563 delT is positively correlated with allele rs17775791 T). The association analysis for tag SNPs was conducted assuming an additive model of genetic effects. For the association analysis of the SABRE cohort, we selected samples of European descent ( $N = 277$ ). We included the first 10 principal components of the kinship matrix to correct the effects of population stratification. Association analysis was performed using PLINK 2.0 (Chang et al. 2015). Association analysis of PainOmics samples was conducted separately for samples of Italian, Belgian, Croatian, and UK descent. Additionally, samples genotyped by GSA and HumanCore SNP arrays were analyzed separately. The first 10 principal components of the kinship matrix were included to correct the effects of population stratification. Association analysis was performed using PLINK 2.0 (Chang et al. 2015). Association analysis of EPIC-Potsdam samples was conducted separately for each batch genotyped in the 3 different platforms. Association analysis of SOCCS samples was performed using SNPTEST v2.5.2 (Marchini et al. 2007). The first 10 principal components of the kinship matrix were included to correct the effects of population stratification. We conducted an inverse-variance weighted fixed-effect meta-analysis of sixteen tag SNPs among four cohorts. The meta-analysis was performed using GWAS-MAP tool (Gorev et al. 2018).

### **Replication**

For the replication we used the protocol, published by (NCI-NHGRI Working Group on Replication in Association Studies et al. 2007). Since there is no direct trait-to-trait correspondence between glycan traits measured by HPLC and UHPLC technologies, we tested the association of tag SNPs with all 117 glycan traits. We considered locus as replicated if the tag SNP showed association with at least one of 117 glycan traits with a replication threshold of  $P \leq 0.05/(16 \times 117) = 2.67 \times 10^{-5}$ , where 16 is the number of tag SNPs and 117 is the number of *N*-glycosylation traits. The summary of the association between 16 loci and glycome traits passed replication threshold is provided in Supplementary Table 2. Statistical power calculation is given in Supplementary Methods.

### **Statistical power calculation**

For each SNP, statistical power (or probability) of replication was estimated using the fact that under alternative hypothesis ( $H_1: \beta \neq 0$ ) the test statistics  $T^2$  from replication sample is expected to follow the  $\chi^2_{df=1, NCP}$  distribution, where NCP is the expected non-centrality parameter computed as  $(T^2 - 1) * N_{rep}/N_{disc}$ , where  $T^2 = (\beta_{disc}/se_{disc})^2$  is test statistic for a particular SNP in the discovery cohort,  $N_{rep}$  is the sample size of the replication cohort and  $N_{disc}$  is the

sample size of the discovery cohort. The power of replication is equal to the probability that statistics following such a distribution would exceed the threshold value  $k = 17.6$  that corresponds to right-hand integral of  $\chi^2_{df=1}$  equal to  $0.05/(117 * 16) = 2.67 \times 10^{-5}$ .

## References

Agakova A, Vučković F, Klarić L, Lauc G, Agakov F. 2017. Automated integration of a UPLC glycomic profile. In: *Methods in Molecular Biology*. Vol. 1503. Humana Press, New York, NY. p. 217–233.

Allegri M, De Gregori M, Minella CE, Klersy C, Wang W, Sim M, Gieger C, Manz J, Pemberton IK, MacDougall J, et al. 2016. ‘Omics’ biomarkers associated with chronic low back pain: protocol of a retrospective longitudinal study. *BMJ Open*. 6(10):e012070. doi:10.1136/bmjopen-2016-012070.

Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT. 2010. Data quality control in genetic case-control association studies. *Nat Protoc*. 5(9):1564–73. doi:10.1038/nprot.2010.116.

Aulchenko YS, Ripke S, Isaacs A, van Duijn CM. 2007. GenABEL: an R library for genome-wide association analysis. *Bioinformatics*. 23(10):1294–1296. doi:10.1093/bioinformatics/btm108.

Benedetti E, Gerstner N, Pučić-Baković M, Keser T, Reiding KR, Ruhaak LR, Štambuk T, Selman MHJ, Rudan I, Polašek O, et al. 2019 Oct 24. Systematic evaluation of normalization methods for glycomics data based on performance of network inference. *bioRxiv*.:814244. doi:10.1101/814244.

Bergmann MM, Bussas U, Boeing H. 1999. Follow-Up Procedures in EPIC-Germany – Data Quality Aspects. *Ann Nutr Metab*. 43(4):225–234. doi:10.1159/000012789.

Boeing H, Korfmann A, Bergmann MM. 1999. Recruitment Procedures of EPIC-Germany. *Ann Nutr Metab*. 43(4):205–215. doi:10.1159/000012787.

Boeing H, Wahrendorf J, Becker N. 1999. EPIC-Germany – A Source for Studies into Diet

and Risk of Chronic Diseases. *Ann Nutr Metab.* 43(4):195–204. doi:10.1159/000012786.

Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience.* 4(1):7. doi:10.1186/s13742-015-0047-8.

Dagostino C, De Gregori M, Gieger C, Manz J, Gudelj I, Lauc G, Divizia L, Wang W, Sim M, Pemberton IK, et al. 2017. Validation of standard operating procedures in a multicenter retrospective study to identify -omics biomarkers for chronic low back pain. Samartzis D, editor. *PLoS One.* 12(5):e0176372. doi:10.1371/journal.pone.0176372.

Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, Vrieze SI, Chew EY, Levy S, McGue M, et al. 2016. Next-generation genotype imputation service and methods. *Nat Genet.* 48(10):1284–1287. doi:10.1038/ng.3656.

Dieterle F, Ross A, Schlotterbeck G, Senn H. 2006. Probabilistic Quotient Normalization as Robust Method to Account for Dilution of Complex Biological Mixtures. Application in 1 H NMR Metabonomics. *Anal Chem.* 78(13):4281–4290. doi:10.1021/ac051632c.

Goldstein JL, Crenshaw A, Carey J, Grant GB, Maguire J, Fromer M, O'Dushlaine C, Moran JL, Chambert K, Stevens C, et al. 2012. zCall: a rare variant caller for array-based genotyping: genetics and population analysis. *Bioinformatics.* 28(19):2543–5. doi:10.1093/bioinformatics/bts479.

Gorev DD, Shashkova TI, Pakhomov E, Torgasheva A, Klaric L, Severinov A, Sharapov S, Alexeev DG, Aulchenko YS. 2018. GWAS-MAP: a platform for storage and analysis of the results of thousands of genome-wide association scans. In: *Bioinformatics of Genome Regulation and Structure\Systems Biology (BGRS\SB-2018).* p. 43.

Grove ML, Yu B, Cochran BJ, Haritunians T, Bis JC, Taylor KD, Hansen M, Borecki IB, Cupples LA, Fornage M, et al. 2013. Best practices and joint calling of the HumanExome BeadChip: the CHARGE Consortium. Aulchenko YS, editor. *PLoS One.* 8(7):e68095. doi:10.1371/journal.pone.0068095.



Guo Y, He J, Zhao S, Wu H, Zhong X, Sheng Q, Samuels DC, Shyr Y, Long J. 2014. Illumina human exome genotyping array clustering and quality control. *Nat Protoc.* 9(11):2643–2662. doi:10.1038/nprot.2014.174.

Huffman JE, Knežević A, Vitart V, Kattla J, Adamczyk B, Novokmet M, Igl W, Pučić M, Zgaga L, Johansson Å, et al. 2011. Polymorphisms in B3GAT1, SLC9A9 and MGAT5 are associated with variation within the human plasma N-glycome of 3533 European adults. *Hum Mol Genet.* 20(24):5000–5011. doi:10.1093/hmg/ddr414.

Johnson WE, Li C, Rabinovic A. 2007. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics.* 8(1):118–127. doi:10.1093/biostatistics/kxj037.

Karssen LC, van Duijn CM, Aulchenko YS. 2016. The GenABEL Project for statistical genomics. *F1000Research.* 5:914. doi:10.12688/f1000research.8733.1.

Langenberg C, Sharp SJ, Franks PW, Scott RA, Deloukas P, Forouhi NG, Froguel P, Groop LC, Hansen T, Palla L, et al. 2014. Gene-Lifestyle Interaction and Type 2 Diabetes: The EPIC InterAct Case-Cohort Study. *PLoS Med.* 11(5). doi:10.1371/journal.pmed.1001647.

Lauc G, Essafi A, Huffman JE, Hayward C, Knežević A, Kattla JJ, Polašek O, Gornik O, Vitart V, Abrahams JL, et al. 2010. Genomics Meets Glycomics—The First GWAS Study of Human N-Glycome Identifies HNF1 $\alpha$  as a Master Regulator of Plasma Protein Fucosylation. Gibson G, editor. *PLoS Genet.* 6(12):e1001256. doi:10.1371/journal.pgen.1001256.

Law PJ, Timofeeva M, Fernandez-Rozadilla C, Broderick P, Studd J, Fernandez-Tajes J, Farrington S, Svinti V, Palles C, Orlando G, et al. 2019. Association analyses identify 31 new risk loci for colorectal cancer susceptibility. *Nat Commun.* 10(1):2154. doi:10.1038/s41467-019-09775-w.

Loh PR, Palamara PF, Price AL. 2016. Fast and accurate long-range phasing in a UK Biobank cohort. *Nat Genet.* 48(7):811–816. doi:10.1038/ng.3571.

Marchini J, Howie B, Myers S, McVean G, Donnelly P. 2007. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet.* 39(7):906–13.

doi:10.1038/ng2088.

McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, Kang HM, Fuchsberger C, Danecek P, Sharp K, et al. 2016. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet.* 48(10):1279–1283. doi:10.1038/ng.3643.

NCI-NHGRI Working Group on Replication in Association Studies, Chanock SJ, Manolio T, Boehnke M, Boerwinkle E, Hunter DJ, Thomas G, Hirschhorn JN, Abecasis G, Altshuler D, et al. 2007. Replicating genotype–phenotype associations. *Nature.* 447(7145):655–660. doi:10.1038/447655a.

Sharapov SZ, Tsepilov YA, Klaric L, Mangino M, Thareja G, Shadrina AS, Simurina M, Dagostino C, Dmitrieva J, Vilaj M, et al. 2019. Defining the genetic control of human blood plasma N-glycome using genome-wide association study. *Hum Mol Genet.* 28(12):2062–2077. doi:10.1093/hmg/ddz054.

Tenesa A, Farrington SM, Prendergast JGD, Porteous ME, Walker M, Haq N, Barnetson RA, Theodoratou E, Cetnarskyj R, Cartwright N, et al. 2008. Genome-wide association scan identifies a colorectal cancer susceptibility locus on 11q23 and replicates risk loci at 8q24 and 18q21. *Nat Genet.* 40(5):631–637. doi:10.1038/ng.133.

Theodoratou E, Thaçi K, Agakov F, Timofeeva MN, Štambuk J, Pučić-Baković M, Vučković F, Orchard P, Agakova A, Din FVN, et al. 2016. Glycosylation of plasma IgG in colorectal cancer prognosis. *Sci Rep.* 6(1):28098. doi:10.1038/srep28098.

Tillin T, Forouhi NG, McKeigue PM, Chaturvedi N, SABRE Study Group. 2012. Southall And Brent REvisited: Cohort profile of SABRE, a UK population-based comparison of cardiovascular disease and diabetes in people of European, Indian Asian and African Caribbean origins. *Int J Epidemiol.* 41(1):33–42. doi:10.1093/ije/dyq175.

Trbojević Akmačić I, Ugrina I, Štambuk J, Gudelj I, Vučković F, Lauc G, Pučić-Baković M. 2015. High-throughput glycomics: Optimization of sample preparation. *Biochem.* 80(7):934–942. doi:10.1134/S0006297915070123.

Vuckovic F, Theodoratou E, Thaci K, Timofeeva M, Vojta A, Stambuk J, Pucic-Bakovic M, Rudd PM, Ereli L, Servis D, et al. 2016. IgG Glycome in Colorectal Cancer. *Clin Cancer Res.* 22(12):3078–3086. doi:10.1158/1078-0432.CCR-15-1867.

Winkler TW, Day FR, Croteau-Chonka DC, Wood AR, Locke AE, Mägi R, Ferreira T, Fall T, Graff M, Justice AE, et al. 2014. Quality control and conduct of genome-wide association meta-analyses. *Nat Protoc.* 9(5):1192–1212. doi:10.1038/nprot.2014.071.