DOI 10.18699/SBB-2022-265

Defining the genetic control of N-glycosylation human blood plasma proteins using multivariate genetic association analysis

Timoshchuk A.[1, 2]*, Potapova N.A.[3], Akmacic I.T.[4], Lauc G.[4], Sharapov S.[1, 3], Aulchenko Y.S.[1, 3]

[1] *MSU Institute for Artificial Intelligence, Lomonosov Moscow State University, Moscow, Russia*

[2] *Moscow Institute of Physics and Technology, Moscow, Russia*

[3] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[4] *Genos Glycoscience Research Laboratory, Zagreb, Croatia*

** timoschukanny@gmail.com*

Key words: human plasma N-glycome, GWAS, MANOVA approach, genetic control

Motivation and Aim: Biochemical pathways underlying protein N-glycosylation are well known, but the biological networks regulating these pathways remain poorly understood. Defining the genetic control of N-glycosylation of proteins of human blood plasma ("plasma N-glycome") can shed light on the regulation of this complex process. To date, a series of genome-wide association studies (GWAS) of human blood plasma protein N-glycosylation have been published reporting more than 30 genetic loci affecting protein glycosylation [1-3]. These studies used a regression model under which glycome traits were analyzed independently of each other. However, it has been previously demonstrated that multivariate genetic association analysis of N-glycome, that is, a joint analysis of multiple glycomic traits, has higher power for loci detection [4, 5]. In this work we aimed to perform the first multivariate GWAS of total plasma N-glycome.

Methods and Algorithms: We utilized MANOVA approach for multivariate analysis of genetic associations. We defined a list of biochemically similar multivariate traits, grouping them on such characteristics as galactosylation and sialylation of antennary branches, core-fucosylation, bisecting GlcNAc, etc. After that, multivariate analysis was carried out using the summary statistics-based method described by Ning et al. [6]. This method enables multivariate analysis of variance using the results of multiple GWAS for individual traits. Moreover, our study included a replication experiment based on the protocol of Ning et al. [6] – reproduction of statistically significant associations in an independent set of samples. For loci that replicated in our study, we performed an *in silico* functional analysis.

Results: We identified 52 loci in total. Compared to results of previous studies we found 17 new loci associated with total plasma N-glycosylation and 7 loci that overlapped with those identified but not replicated previously. We replicated 8 loci and compiled a list of candidate genes for each of them.

Conclusion: Identification of new loci associated with total plasma N-glycosylation can provide new knowledge towards understanding the mechanisms of *in vivo* regulation of this process. Among the genes prioritized as possible candidates, several could be linked to N-glycosylation and/or immune function.

Acknowledgements: The work of NP, SS and YA was partly supported by a grant from the Russian Science Foundation (RSF) No. 19-15-00115.

References

1. Lauc G. et al. The First GWAS Study of Human N-Glycome Identifies HNF1α as a Master Regulator of Plasma Protein Fucosylation. *PLoS Genet*. 2010;6(12):e1001256.
2. Huffman J.E. et al. Polymorphisms in B3GAT1, SLC9A9 and MGAT5 are associated with variation within the human plasma N-glycome of 3533 European adults. *Hum Mol Genet*. 2011;20(24):5000-5011.
3. Sharapov S.Z. et al. Defining the genetic control of human blood plasma N-glycome using genome-wide association study. *Hum Mol Genet*. 2019;28(12):2062-2077.
4. Shen X. et al. Multivariate discovery and replication of five novel loci associated with Immunoglobulin G N-glycosylation. *Nat Commun*. 2017;8(1):1-10.
5. Shadrina A.S. et al. Multivariate genome-wide analysis of immunoglobulin G N-glycosylation identifies new loci pleiotropic with immune function. *Hum Mol Genet*. 2021;30(13):1259-1270.
6. Ning Z. et al. Nontrivial replication of loci detected by multi-trait methods. *Front Genet*. 2021;12:627989. doi: 10.3389/fgene.2021.627989.