DOI 10.18699/SBB-2022-213

# Re-analysis of DNA methylation data from tuberculosis-infected individuals with a focus on target genes

Babushkina N.P.*, Bragina E.Yu., Goncharova I.A., Markov A.V., Kucher A.N., Nazarenko M.S.

*Research Institute of Medical Genetics, TNRMC RAS, Tomsk, Russia*

** nad.babushkina@medgenetics.ru*

**Key words:** tuberculosis, methylation

*Motivation and Aim*: Along with the structural variability of the genome, epigenetic modifications can be significant for susceptibility to various, including infectious diseases. The aim of this study was to re-analyze the differential methylation of genes (DMG), for which associations with tuberculosis and its pathogenetically significant features were previously established, as well as for those of interest from the point of view of studying the phenomenon of reverse comorbidity of asthma and tuberculosis.

*Methods and Algorithms*: Re-analysis of two DMG datasets presented in the public repository of functional genomic data Gene Expression Omnibus [1] was carried out. The GSE72338 dataset [2] was obtained from the study of DNA methylome, transcriptome (mRNA and microRNA), and proteome. The study was performed on peripheral blood monocytes (PBMC) and neutrophils of patients with active tuberculosis (TB, $n = 8$) and clinically healthy individuals, with household contact with active TB patients (LTBI, $n = 8$)). The patients lived in an area with a high incidence TB in Cape Town, South Africa [2]. The second dataset, GSE118469 [3], was obtained from a study of TB in Taiwanese residents. The study included patients with TB ($n = 12$) without concomitant infections (including HIV) and malignant neoplasms. The control group comprised of healthy individuals ($n = 6$). Methylation analysis was performed on an Infinium HumanMethylation450K BeadChip v1.2 chip [3].

In the GSE72338 and GSE118469 datasets we were primarily interested in genes that, according to our previously data were associated with infectious diseases, in particular thirteen genes *IFNG, SOCS5, TNFB, TNFRSF1B, PIAS3, PIASY, CXCL10, ATM, NBN, MRE11, MLH1, PMS2,* and *TP53BP1*.

*Results*: Only the *SOCS5* gene (out of 13 genes) included in the Top-50 DMG list, topped the DMG list when comparing neutrophils from TB and LTBI patients ($p = 0.00000226$), but adjusting for multiple comparisons makes these differences not statistically significant (adj.$p = 0.595$). It should be noted that, taking into account the FDR correction no significant differences were obtained in any of the comparisons. Further, we carried out the analysis, focusing on the initial level of significance achieved, without introducing any corrections.

In monocytes, 13 sites in seven target genes (*NBN, MRE11A, MLH1, IFNG, LTA, TNFRSF1B, CXCL10*) were differentially methylated in patients with active TB compared with LTBI. In neutrophils, 7 sites in regions of genes *ATM, MLH1, PMS2, SOCS5, TNFRSF1B, CXCL10* were differentially methylated in patients with active TB compared with LTBI. 23 sites in *ATM, PMS2, TP53BP1, SOCS5, TNFRSF1B, CXCL10* genes were differentially methylated in PBMC when comparing between TB patients and healthy individuals.

Then, we annotated these loci with the UCSC resource. In two datasets, 39 CpG-sites were identified in target genes that were differentially methylated between TB patients and healthy controls. The position of three CpG-sites in the genome allows them to be attributed to the sphere of influence of two genes: cg11588932 (*ATM, NPAT*), cg13846866 (*MLH1, EPM2AIP1*), cg23492613 (*MRE11A, ANKRD49*). There are 32 sites in the gene body (targeted or co-localized) (4 in exons and 28 in introns); 7 sites are located in the intergenic space near the promoter region of the gene.

The methylation level of most of the studied sites (> 200 or even > 600) indicates their low transcriptional activity. This is supported by a small number (8 CpG) sites localized in the RNA polymerase II subunit A binding region and in open heterochromatin (12 CpG sites).The CpG-islands include 5 sites, and 21 sites are located near the islands, 12 sites are single. At that, 26 sites are localized near the promoters of target (and co-localized) genes; only cg21288207 is included directly in the promoter of the target gene (*PMS2*). In addition, 2 CpG-sites are included in the co-localized genes promoter sequence (cg23492613: *ANKRD49*; cg16280132: *ATP6V1G2*). Only 9 CpG-sites are located in transcription factor binding regions. However, cg24682307 (*TNFRSF1B*) is located in the IRF1 binding region, which activates the transcription of genes involved in the response to viruses and bacteria [4]; cg21288207 and cg19570749 (*PMS2*) in the ELF1 and GATA3 binding regions respectively, that regulate gene expression during the development of T-lymphocytes [5, 6].

*Conclusion*: Thus, these findings indicated that DNA methylation is important in TB development. The differential value of DNA methylation of 13 target genes associated with TB, as well as the localization of these sites in the binding regions of transcription factors involved in the development of the immune response to infection, confirm the involvement of protein products of target genes in the pathogenesis of TB.

*Acknowledgements*: The work was carried out within the framework of the State Task of the Ministry of Science and Higher Education No. 122020300041-7.

*References*

1. GEO, https://www.ncbi.nlm.nih.gov/geo.
2. Esterhuyse M.M. et al. Epigenetics and proteomics join transcriptomics in the quest for tuberculosis biomarkers. *mBio.* 2015;6(5):e01187-e01115. doi: 10.1128/mBio.01187-15.
3. Chen Y.C. et al. Whole genome DNA methylation analysis of active pulmonary tuberculosis disease identifies novel epigenotypes: PARP9/miR-505/RASGRP4/GNG12 gene methylation and clinical phenotypes. *Int J Mol Sci*. 2020;21(9):3180. doi: 10.3390/ijms21093180.
4. Taki S. et al. Multistage regulation of Th1-type immune responses by the transcription factor IRF-1. *Immunity*. 1997;6(6):673-679. doi: 10.1016/s1074-7613(00)80443-4.
5. Leiden J.M. et al. A novel Ets-related transcription factor, Elf-1, binds to human immunodeficiency virus type 2 regulatory elements that are required for inducible trans activation in T cells. *J Virol.* 1992;66(10):5890-5897. doi: 10.1128/JVI.66.10.5890-5897.1992.
6. Wan Y.Y. GATA3: a master of many trades in immune regulation. *Trends Immunol.* 2014;35(6):233-242. doi: 10.1016/j.it.2014.04.002.