doi 10.18699/bgrs2024-1.4-31

Candidate SNP markers significantly altering the affinity of TATA-binding protein for the promoters of human hub genes for atherogenesis, atherosclerosis and atheroprotection

Tverdokhleb N.[1]*, Chadaeva I.[1], Filonov S.[1], Zolotareva K.[1], Podkolodnyy N.[1, 3], Khandaev B.[1, 2], Rasskazov D.[1], Bogomolov A.[1, 2], Kazachek A.[1, 2], Oshchepkov D.[1, 2], Ivanisenko V.[1, 2], Demenkov P.[1], Podkolodnaya O.[1], Ponomarenko P.[1], Mustafin Z.[1, 2], Kondratyuk E.[1, 4, 5], Savinkova L.[1], Ponomarenko M.[1], Suslov V.[1], Kolchanov N.[1, 2]

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

[2] *Novosibirsk State University, Novosibirsk, Russia*

[3] *Institute of Computational Mathematics and Mathematical Geophysics, SB RAS, Novosibirsk, Russia*

[4] *Research Institute of Clinical and Experimental Lymphology – Branch of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[5] *Siberian Federal Scientific Centre of Agro-BioTechnologies, RAS, Krasnoobsk, Novosibirsk region, Russia*

* *nata@bionet.nsc.ru*

Key words: atherogenesis; atheroprotection; atherosclerosis; hub gene; promoter; TBP; TATA box; SNP; candidate SNP marker; gene expression change; natural selection; *in silico* verification

Motivation and Aim: Atherosclerosis is a systemic disease in which focal lesions in arteries promote the build-up of lipoproteins and cholesterol they are transporting. The development of atheroma (atherogenesis) narrows blood vessels, reduces the blood supply and leads to cardiovascular diseases. According to the World Health Organization (WHO), cardiovascular diseases are the leading cause of death, which has been especially boosted since the COVID-19 pandemic. There is a variety of contributors to atherosclerosis, including lifestyle factors and genetic predisposition. Antioxidant diets and recreational exercises act as atheroprotectors and can retard atherogenesis. The search for molecular markers of atherogenesis and atheroprotection for predictive, preventive and personalized medicine appears to be the most promising direction for the study of atherosclerosis.

Methods and Algorithms: Here, we have analyzed 1068 human genes associated with atherogenesis, atherosclerosis and atheroprotection [1], as depicted in the Figure 1.

Results: The hub genes regulating these processes have been found to be the most ancient by means of the an *in silico* phylostratigraphic analysis [2, 3]. *In silico* analysis using our freely available web service SNP_TATA_Comparator [4] of all 5112 SNPs in their promoters has revealed 330 candidate SNP markers, which statistically significantly change the affinity of the TATA-binding protein (TBP) for these promoters, as shown in the Table 1.

Conclusion: These molecular markers have made us confident that natural selection acts against underexpression of the hub genes for atherogenesis, atherosclerosis and atheroprotection, while upregulation of the only atheroprotection-related genes promotes human health according to the Table 2.

Funding: The study is supported by the Government Budget Project FWNR-2022-0020.

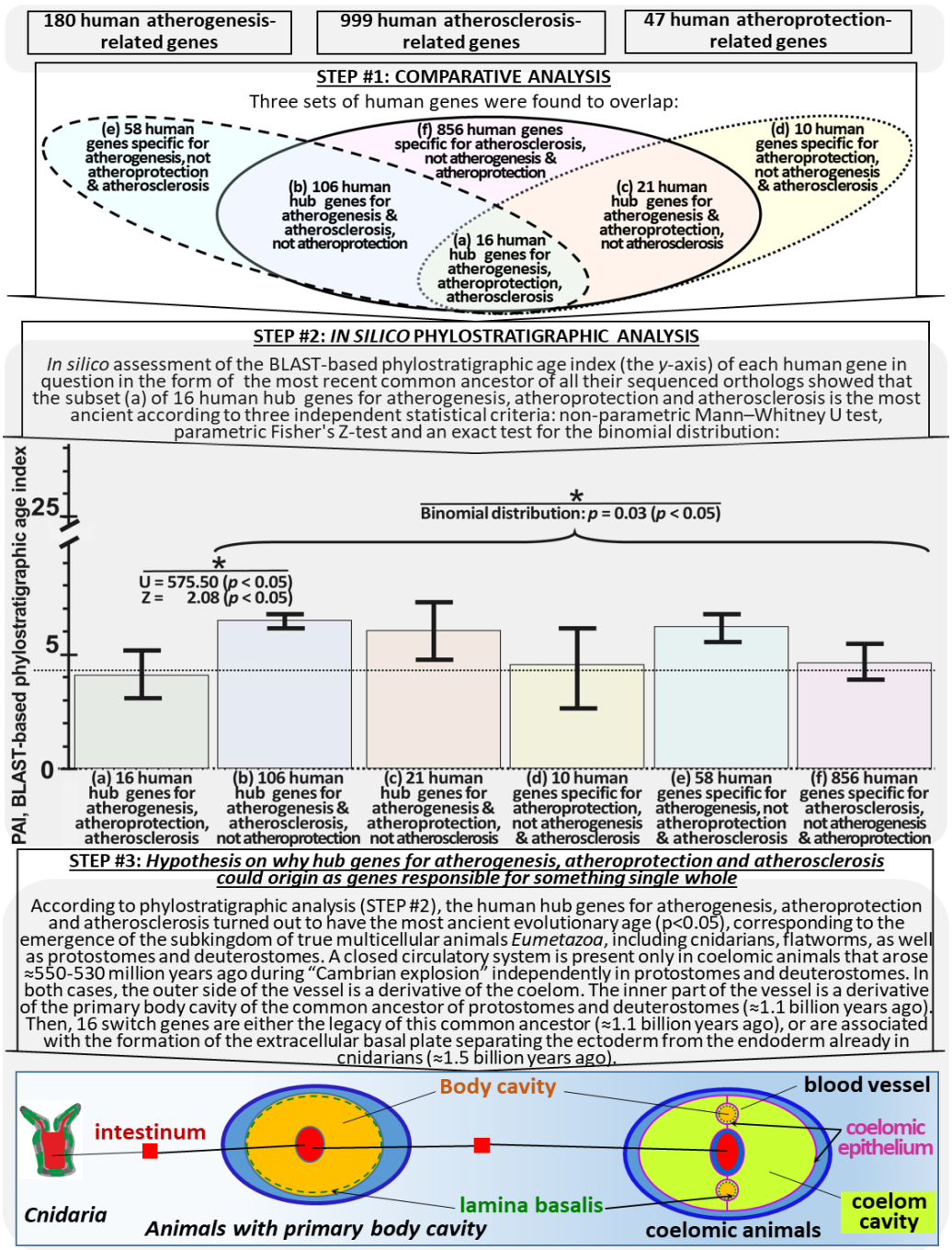

Fig. 1. The flowchart depicting step-by-step how we studied all 1068 human genes associated with atherogenesis, atherosclerosis and atheroprotection according to the NCBI Gene database [1]. *Legend*: PAI, genes' phylostratigraphic age index evaluated against the BLAST-based scale [2] using the freely available web service Orthoscape [3]. the heights of the gene bars and error bars, the arithmetic mean of the PAIs and the standard error of the mean (SEM); the dotted line, an auxiliary line running above the gray-green bar for the subset (a) of 16 human hub genes for atherogenesis, atherosclerosis and atheroprotection, the tops of the other five bars being above the line. (*): statistical significance $p < 0.05$. U and Z are the statistics in the nonparametric Mann–Whitney U test and parametric Fischer's Z-test according to Statistica (StatSoft™, USA)

Table 1. Candidate SNP makers significantly upregulating and downregulating TATA binding protein (TBP) affinity for the promoters of the human hub genes for atherogenesis, atherosclerosis and atheroprotection found in this work in comparison with those independently identified within whole human genome [5, 6] as an indicator of the neutral drift according to Haldane's dilemma [7] and neutral evolution theory [8]

Human Genome Assembly GRCh38/hg38, dbSNP Rel. 155	Number of objects ($N_{OBJECTS}$; "↑", upregulation; "↓", downregulationt)					Neutral drift: $p(H_0: N_{\downarrow} \geq 4N_{\uparrow})$ [7, 8]
	N_{GENE}	N_{SNP}	N_{MARKER}	$N_{\downarrow}$	$N_{\uparrow}$	Binomial distribution, *Bonferroni's correction*, P_{ADJ}
Whole-genome norm for SNPs significantly altering TBP-sites [5, 6]	30,000	10^5	1000	800	200	1.00
Human hub genes for athero-genesis, atheroprotection, and atherosclerosis [this work]	16	5112	466	150	316	10^{-4}

Table 2. Candidate SNP makers significantly changing TBP affinity for the promoters of the human hub genes for atherogenesis, atherosclerosis and atheroprotection

Atherosclerosis-related processes	The number of candidate SNP markers that have been found to significantly change TBP affinity for the promoters of the hub genes for atherogenesis, atherosclerosis and atheroprotection		Statistical significance binomial distribution, Bonferroni's correction, P_{ADJ}
	Human health failure	Human health improvement	
Atherogenesis	186	144	1.00
Atherosclerosis	195	135	0.19
Atheroprotection	**91**	**239**	**10^{-4}**

References

1. Brown G.R., Hem V., Katz K.S., Ovetsky M., Wallin C., Ermolaeva O., Tolstoy I., Tatusova T., Pruitt K.D., Maglott D.R., Murphy T.D. Gene: a gene-centered information resource at NCBI. *Nucleic Acids Res*. 2015;43:D36-D42. doi 10.1093/nar/gku1055
2. Altschul S.F., Gish W., Miller W., Myers E.W., Lipman D.J. Basic local alignment search tool. *J Mol Biol*. 1990;215(3):403-410. doi 10.1016/S0022-2836(05)80360-2
3. Mustafin Z.S., Lashin S.A., Matushkin Y.G., Gunbin K.V., Afonnikov D.A. Orthoscape: a cytoscape application for grouping and visualization KEGG based gene networks by taxonomy and homology principles. *BMC Bioinformatics*. 2017;18(Suppl. 1):1427. doi 10.1186/s12859-016-1427-5
4. Ponomarenko M., Rasskazov D., Arkova O., Ponomarenko P., Suslov V., Savinkova L., Kolchanov N. How to use SNP_TATA_Comparator to find a significant change in gene expression caused by the regulatory SNP of this gene's promoter via a change in affinity of the TATA-binding protein for this promoter. *Biomed Res Int*. 2015;2015:359835. doi 10.1155/2015/359835
5. 1000 Genomes Project Consortium; Abecasis G.R., Auton A., Brooks L.D., DePristo M.A., Durbin R.M., Handsaker R.E., Kang H.M., Marth G.T., McVean G.A. An integrated map of genetic variation from 1,092 human genomes. *Nature*. 2012;491(7422):56-65. doi 10.1038/nature11632
6. Kasowski M., Grubert F., Heffelfinger C., Hariharan M., Asabere A. et al. Variation in transcription factor binding among humans. *Science*. 2010;328(5975):232-235. doi 10.1126/science.1183621
7. Haldane J.B.S. The cost of natural selection. *J Genet*. 1957;55:511-524
8. Kimura M. Evolutionary rate at the molecular level. *Nature*. 1968;217(5129):624-626. doi 10.1038/217624a0