doi 10.18699/bgrs2024-1.4-24

Ontology analysis of big transcriptomic data and differential gene expression

Oshchepkov D.[1], Chadaeva I.[1]*, Kozhemyakina R.[1], Zolotareva K.[1], Khandaev B.[1], Sharypova E.[1], Ponomarenko P.[1], Bogomolov A.[1], Klimova N.V.[1], Shikhevich S.[1], Redina O.[1], Kolosova N.G.[1], Nazarenko M.[2], Kolchanov N.A.[1], Markel A.[1], Ponomarenko M.[1]

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

[2] *Institute of Medical Genetics, Tomsk National Research Medical Center, Tomsk, Russia*

* *ichadaeva@bionet.nsc.ru*

Key words: hypertension; stress reactivity; molecular marker; age-related disease; differentially expressed gene; functional annotation of genes; gene network; correlation; principal component; bootstrap

Motivation and Aim: The aim of this study was to conduct an ontological analysis of the transcriptome of two lines of gray rats, tame and aggressive, as well as identified differentially expressed genes, and compare them with human transcriptomic data in accordance with the phenotype of a specific disease.
Methods and Algorithms: To analyze the DEGs, we used our previously obtained high-throughput RNA sequencing data from four brain regions of gray rats of the two strains: the hypothalamus [1], the hippocampus [2], the periaqueductal gray (PAG) [3], and the midbrain tegmentum (MTg) [4]. DEGs were subjected to principal component analysis using the PAST4.04 software package. We used available transcriptome data from tissues from patients with diseases compared with apparently healthy controls, model animals compared with controls, and domestic animals compared with wild animals. Using the PubMed database, we characterized rat DEGs and human DEGs in terms of the clinical implications of their under- or over-expression for disease. All these data were statistically significant according to Fisher's Z test, adjusted for multiple comparisons.
Results: The hippocampal transcriptome of tame and aggressive gray rats was analyzed using an ontological approach to investigate molecular markers of hypertension. It is known that increased stress reactivity increases the risk of developing arterial hypertension. In this regard, an analysis of the identified DEGs was carried out in the ontological data space defined by the concepts “Gene Expression”, “Stress Reactivity/Domestication” and “Hypertension”. This approach collected all transcriptomic data from animal models of hypertension available at the time of the study. These are 14 models for 4 animal species in which hypertension was studied: rats, mice, rabbits and chickens, for a total of 4216 DEGs. We also collected all available transcriptomic data from studies of patients with hypertension compared with apparently healthy controls, for which a total of 7865 DEGs were identified (Figure 1). In total, these two samples amounted to 12081 DEGs. Thus, an analysis that started with 42 DEGs resulted in the need to consider a big data sample that was 300 times larger in size. Analysis in each of the samples revealed a significant correlation between changes in expression levels obtained in experiments with stress reactivity in tame and aggressive

rats, as well as in experiments with hypertension for homologous genes in model animals and humans.

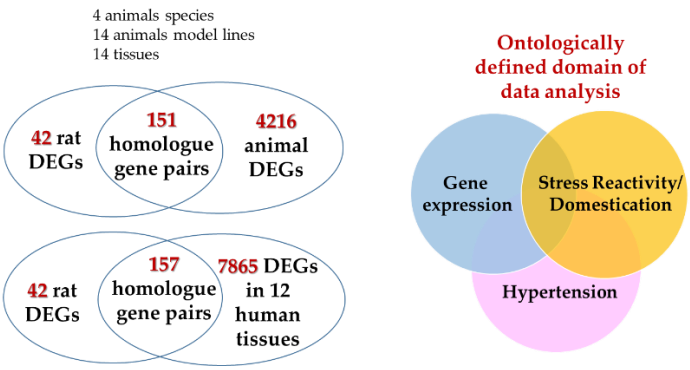

Fig. 1. Flowchart for ontological analysis of big transcriptomic data and differentially expressed genes in association with stress reactivity and hypertension

Analysis of these dependencies by the principal component method made it possible to obtain expressions for two components, one of which corresponded to arterial hypertension, and the second to stress reactivity. At the last stage, for each of the identified 42 DEGs in the hippocampus of tame and aggressive rats and the corresponding homologous human DEGs, the number of homologous pairs of DEGs with the same signs of the principal components, calculated by the $\log_2$ value of expression changes and their statistical significance, was determined. Identification of genes correlated with PC1 has identified two molecular markers associated with the risk of developing stress-sensitive hypertension: the β-protocadherin gene *Pcdhb9* (overexpression of which narrows the internal diameter of blood vessels) and the hemoglobin β-subunit gene, *Hbb-b1* (overexpression of which increases viscosity blood) [2].

We then focused on PAG to investigate molecular markers of age-related diseases (ARD). A large sample of different domestic animals shows an accumulation of genetic mutations that contribute to susceptibility to such diseases. Similar ontological analysis and principal component analysis of a big dataset identified only one gene out of 39 DEGs as a molecular marker associated with the risk of developing ARD. In particular, the *Fcgr2b* gene in the PAG of tame and aggressive rats correlated with susceptibility to ARD (Table 1). Increased expression of this gene aggravates immunogenic diseases associated with cell aging [3].

Table 1. Comparison of the effects of unidirectional changes in the expression of the human *FCGR2B* gene on age-related diseases (ARD) severity in humans and of its animal homologs on the microevolutionary events leading to domestic and wild animals

Effect of changes in human FCGR2B expression on ARD		Effect of changes in the expression of animal genes homologous to human FCGR2B					
Decreased expression	Increased expression	Decreased expression	Increased expression	Tissue	DEG	$\log_2$	P_{adj}
Aggravating age-related diseases: In South Asia and Africa, the human *FCGR2B*-deficient alleles have the most frequent occurrence as protectors against infection, susceptibility to which increases with age as immuno-senescence	Alleviating age-related diseases: *FCGR2B* has been explored using the "C-type lectin-like molecule-1"/"Fc-domain" fusion protein as a target antigen for chemotherapy against acute myeloid leukemia as a cellular-senescence-related immunogenic disease	Aggressive rat	Tame rat	Periaqueductal gray	*Fcgr2b*	2.02	0.05
				Hypothalamus	*Fcrl2*	1.12	0.05
					Fcgr3a	2.06	10^{-2}
				Midbrain Tegmentum	*Fcgr2b*	2.01	0.05
		Aggressive fox	Tame fox	Pituitary	*Fcrl1*	0.43	10^{-2}
		Wild rabbit	Domestic rabbit	Parietal-temporal cortex	*Fcgr3b*	1.35	10^{-2}

The transcriptomes of domestic and wild animals were processed accordingly and compared with the transcriptome of the hypothalamus of tame and aggressive gray rats. Here we also identified a number of genes correlated with principal components that reflect the effects of artificial selection and intraspecific variation [1].

Finally, we analyzed the 4th transcriptome, MTg, in the light of the influence of transcription factors on gene expression. Using bioinformatics methods, the correspondence between the increased level of expression of the *Ascl3* gene in tame rats and changes in the expression of genes from the list of DEGs in the MTg was analyzed. It was shown that motifs corresponding to potential binding sites for the transcription factor ASCL3 constitute the majority of all significantly enriched motifs for genes with reduced expression in tame animals. These results led us to the conclusion that the period of neurogenesis in tame rats is more extended than in aggressive ones.

Conclusion: Using an ontological approach and principal component analysis, we identified molecular markers of hypertension and age-related diseases, and also identified genes associated with domestication. In addition, it was shown that motifs corresponding to potential binding sites for the transcription factor ASCL3 constitute a significant part of all significantly enriched motifs for genes with reduced expression in tame animals, which indicates prolonged neurogenesis.

Funding: The study is supported by Russian government projects FWNR-2022-0015 and Kurchatov Genomic Center of ICG SB RAS agreement number 075-15-2019-1662. For data analysis, computational resources of the Bioinformatics CPC were used with the support of budget project No. FWNR-2022-0020.

References

1. Chadaeva et al. Domestication explains two-thirds of differential gene expression variance between domestic and wild animals; the remaining one-third reflects intraspecific and interspecific variation. *Animals*. 2021;11:2667. doi 10.3390/ani11092667
2. Oshchepkov et al. Stress reactivity, susceptibility to hypertension, and differential expression of genes in hypertensive compared to normotensive patients. *Int J Mol Sci.* 2022;23(5):2835. doi 10.3390/ijms23052835
3. Shikhevich et al. Differentially expressed genes and molecular susceptibility to human age-related diseases. *Int J Mol Sci*. 2023;24(4):3996. doi 10.3390/ijms24043996
4. Oshchepkov et al. Transcription factors as important regulators of changes in behavior through domestication of gray rats: quantitative data from RNA sequencing. *Int J Mol Sci.* 2022;23(20):12269. doi 10.3390/ijms232012269