Analysis of differential gene expression by RNA-seq data in brain areas of laboratory animals

Vladimir N. Babenko[1,2], Anatoly O. Bragin[1], Anastasia M. Spitsina[2], Irina V. Chadaeva[1], Elvira R. Galieva[1], Galina V. Orlova[1], Irina V. Medvedeva[2], Yuriy L. Orlov[1,2 *],

[1]Institute of Cytology and Genetics SB RAS, 630090 Novosibirsk, Russia, www.bionet.nsc.ru
[2]Novosibirsk State University, 630090 Novosibirsk, Russia, lcg.nsu.ru

Summary

Computer analysis of gene expression in the nervous system plays a fundamental role in biology, genetics, and neurosciences. We studied molecular and genetic mechanisms of enhanced aggressiveness in comparison with tolerant behaviour using experimental animal models developed at the Institute of Cytology and Genetics SB RAS. Grey rats (*Rattus norvegicus*) have been subjected to selection during several generations in two directions – friendly, tolerant behaviour towards man (tame grey rats) and increased aggressive behaviour. We used samples from hypothalamus, mesencephalic tegmentum and periaqueductum grey matter from brain areas of grey rats genetically selected by behaviour in many generations. The set of computer tools and data processing pipelines helped to find genes and gene regulation patterns related to behaviour patterns. RNA-profiling experiments revealed the lists of differentially expressed genes in the contrast samples as well as differentially spliced isoforms. The gene ontology categories of protein transport, phosphoproteins, and nucleotide binding are presented together with categories of transmission of nerve impulses and neuron development were identified. Differential alternative splicing events found in the brain areas studied are statistically significant. We discuss role of alternative splicing events for neurospecific genes in behaviour patterns as well as extension of brain transcriptomics profiling.

1 Introduction

Aggressive behaviour is a complex phenomenon implying genetics and physiological roots [1]. Basic studies have shown that the frequency and the severity of aggression rate depend on the hereditary predispositions, previous experience of aggressive behaviour and social context, provoking the demonstration of aggression. To study a genetic component of aggressive behaviour we used published data on genes expression (RNA-seq and microarrays) related to such behaviour in mouse and rat, as well as in-house experimental data in rat [2-5].

Computer analysis of molecular mechanisms underlying nervous system function has a fundamental role in biology, genetics, and cognitive sciences [6]. Computer methods are widely used for the study of gene structure, gene interactions, and gene expression regulation in the brain cells. High-throughput sequencing technologies and microarray assays permit us to conduct research on gene expression in brain areas at new, qualitatively higher level. Affymetrix microarrays were used to detect differences in brain gene expression between two inbred mouse strains (C57BL/6J and 129SvEv) [5]. In this work we briefly review the main bioinformatics approaches and databases on gene expression analysis in brain cells and present results on alternative splicing analysis in the rat transcriptomes studied.

* To whom correspondence should be addressed. Email: orlov@bionjet.nsc.ru

doi:10.2390/biecoll-jib-2016-292

The research was aimed to study the molecular and genetic mechanisms of enhanced aggressiveness in comparison with tolerant behaviour using two unique experimental models developed at the Institute of Cytology and Genetics SB RAS. One of them are grey rats (*Rattus norvegicus*), which have been subjected to selection during several generations in two directions – friendly, tolerant behaviour towards a man (tame grey rats) and increased aggressive behaviour ("aggressors") [2]. The latter rats feature the reinforced enhanced aggression not only towards a man (in the "glove test" when the rat in cage attacks and bites human hand in glove), but also towards the animals in the intermale agonistic interactions.

Another behaviour model in mouse checks the increased aggressiveness as the result of repeated positive fighting experience in daily encounters that simulates specific social "experience" [3-5]. As a consequence of the repeated aggression, the social and individual behaviours become pathological: such males attacked females or other mice demonstrating submissive behaviour. The research was aimed to elucidate the genetic and molecular mechanisms of hereditary defined (first model, rats) and acquired (second model, mouse) increased aggressiveness using gene expression profiling by RNA-seq in different brain regions from aggressive and tame animals.

There are several studies of gene expression in brain areas of laboratory animals aimed at general mapping of transcription activity [6-8]. Some differentially expressed genes were found in chromosomal regions with known behavioural quantitative trait loci (QTLs). Our study proceeds with recent published work on group comparison of RNA-seq profiles in tame and aggressive laboratory animals [4; 5]. Integration of available genomics data is an important step for the behavioural studies in rat [1; 2]. We used several computer tools for construction of differentially expressed genes in rat transcriptome sets, as well as new ExpGene tool for analysis of gene co-expression patterns. The gene ontology categories of protein transport, phosphoproteins, as well as transmission of nerve impulses and neuron development categories, were identified as significantly abundant in the differentially expressed gene lists related to aggressive behaviour. In addition, using the software for gene splicing analysis we found the differentially spliced gene isoforms frequency profiles in our RNA-Seq data in aggressive and tame rats. Differential alternative splicing events studied are statistically significant. Finally, we discuss role of the alternative splicing for neurospecific genes in behaviour patterns.

2 Materials

Selection of brain areas responsible for behavioural traits is a key for cell sample preparation. The status of gene expression at the level of mRNA transcription is vital for both neuron development and morphogenesis, as well as for the functioning of distinct gene networks in the mature cells of various brain regions. Transcriptomes obtained from different brain regions with different specializations in aggressive animals can shed light on mechanisms governing the selective functions of these regions and allow the evaluation of their specificity in comparison to nonaggressive animals regardless of whether the animals were the result of selection or the environment. The brain regions in our studies include the mesencephalic tegmentum (MT), containing bodies of dopaminergic and opioidergic neurons and responsible for reward mechanisms; the periaqueductum grey matter (PGM), containing the bodies of serotonergic neurons responsible for the inhibiting control of aggressive behaviour; and the hypothalamus (Hyp), which is involved in the regulation of stress responsiveness.

The elucidation of differentially expressed genes associated with the specific phenotype can be further examined using a variety of secondary analyses, e.g., examining if the genes cluster within known gene ontology categories or are part of a known protein-protein interaction network [6; 9-10]. Selected genes can be grouped on the basis of common transcription factors and other regulatory elements.

3 Methods

RNA-seq sequencing of rat and mouse brain areas samples was done using Illumina HiSeq 1500. RNA quality was tested on Bioanalyzer 2100 (Agilent) following the Illumina protocol. Sequencing depth consisted of at least 10 mol reads for each sample. The files obtained in "fastq" format were used for the reads mapping onto the RGSC Rnor_5.0\rn5 and GRCm38 reference genomes using Tophat2 aligner. Input reads were filtered and trimmed by Trimmomatic software [11]. Trimmomatic omits technical sequences and sequences with low quality from input data. We use Tophat2 program [12] for mapping reads on the rodent reference genomes. Tophat2 is built on ultrafast, memory-efficient program Bowtie [13] and can identify splice junctions. Files with genome annotations were downloaded from UCSC genome browser [14]. The bam-files - received from Tophat output - were used for detection of differentially expressed genes and alternative splicing analysis in the samples. We have analysed rat and mouse gene expression by Cufflinks v2.0.2 programs [10]. Cufflinks use bam files - with mapped reads – and gff files with reference genome annotation for calculation gene expression. Cufflinks also provide information about differently expressed genes between the samples. Gene expression levels were estimated in FPKM (fragments per kilobase per million reads) values. The detection of splice variants and differential splicing was performed using rMATS software [15].

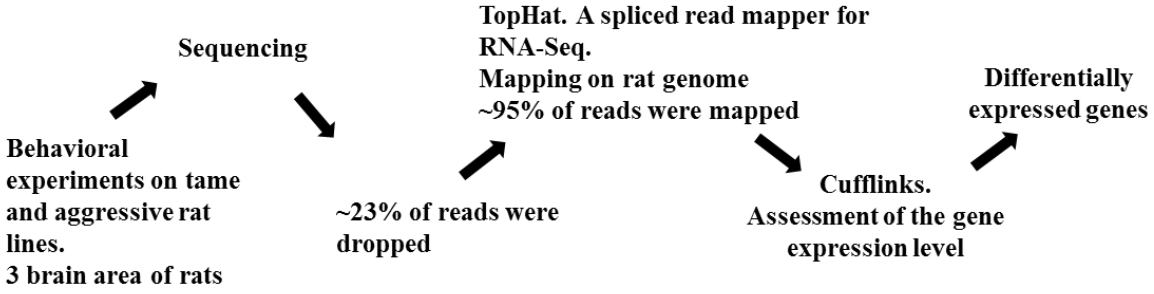

Figure 1: Block scheme for analysis of differentially expressed genes in rat.

4 Software and applications

Earlier we implemented a simple and fast program tool in C++ language for similar tasks of microarray and RNA-Seq data analysis. The tool simplifies the identification of structural features of genes with high expression [16]. It continues approaches developed earlier at ICGenomics complex [17] and software for genome sequences analysis [18]. The ExGene program has user interface, new options for data pre-processing, statistical analysis and data visualization (Figure 2).

Data pre-processing and filtering were necessary because of the technical errors and the redundancy on microarrays, which can lead to incorrect results [19]. Such redundancy often related to low complexity regions in sequences that could be detected by DNA complexity measures implemented in the Complexity program [20, 21]. Note that problem is rather complex and includes entropy estimates in different forms [22]. Filtering of next-generation sequencing data based on Illumina platform and statistical estimates was suggested in [23]. The current version of the software suite includes options for Affymetrix microarray filtering from repeated probes.

The statistical analysis options include calculating of correlation coefficients for two types of coefficients – Pearson linear correlation and Spearman rank correlation. Thus, we follow approaches for gene co-expression analysis presented in [9]. Visualization in our software includes options for plotting expression distribution in bar graph, gene network, and tissue-specificity graph construction for a given gene list. Bar graph plotting on the set of correlation coefficients shows coefficients distribution. Gene network can be visualized using a JavaScript application and represents the graph where the vertices correspond to genes.

We performed a comparative analysis of expression data for genes whose expression is increased in the brain tissue based on BioGPS data (http://biogps.org) and then analysed the expression pattern of pairs of transcripts, co-localized in the genome.

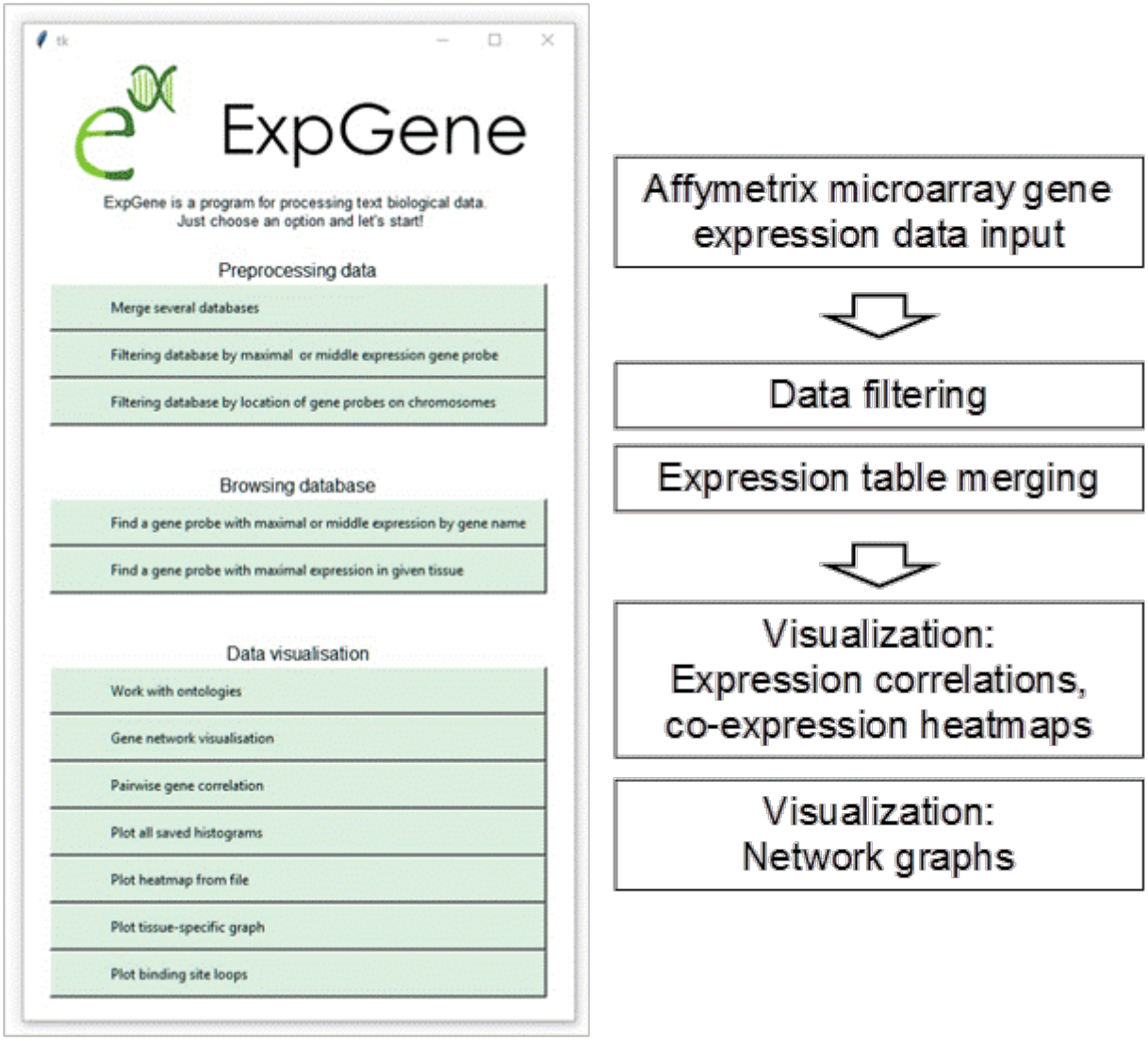

Figure 2: Interface of ExpGene program for gene expression and data workflow analysis (left panel) and microarray data processing for network analysis.

To test the developed software, we considered the correlation of genes expression in the gene networks of circadian rhythm and cholesterol regulation, annotated in GeneNet database [24] as well as the genes which are reportedly responsible for the aggressive behaviour in mice [12; 25].

In particular, we studied the gene networks structure formed by genes which are related to the aggressive behavior using Internet-available information resources GeneNet (http://wwwmgs.bionet.nsc.ru/mgs/gnw/), BioGPS (http://biogps.org) [26; 27]. After constructing the lists of differentially expressed genes we used a set of tools for co-expression analysis to elucidate the characteristic features of gene network, which may be related to aggressive behavior. Standard approaches here is reconstruction of gene networks from a gene list using manual curation such as GeneNet [24], annotation of protein-protein interactions - STRING tool (Search Tool for the Retrieval of Interacting Genes) [28] (http://string-db.org/) and GeneMANIA (Multiple Association Network Integration Algorithm) [29; 30]. The bioinformatics tools such as GeneNet [24], ANDVisio and ANDSystem [31; 32] allow to reconstruct gene networks, corresponding to assemblies of coordinately functioning genes controlling for biochemical, molecular-genetic, and physiological processes based on the published data. Moreover, the complex gene networks associated with behavior and physiological functions of an organism could be reconstructed by analysis of gene associations, mentioned in science literature and other published sources [31; 32]. Thus, direct association of a disease or behavior term in science text could give background for associative network reconstruction using ANDSystem tool [32; 33].

RNA sequencing (RNA-Seq) has become a powerful approach not only for gene expression level analysis, but for genome-wide analysis of pre-mRNA alternative splicing events. To analyze alternative splicing events in the transcriptomics data MATS (multivariate analysis of transcript splicing) and rMATS (replicate MATS) shown to be effective tools [14; 33]. We implemented a Python-based application to tackle rMATs annotation. The rMATs output lacks the information on the short isoforms spectra of particular exon. We implemented the add-on that, based on 3 subsequent exons locations used in alternative exons identification, retrieves from RefGene database short isoforms IDs lacking the exon. While it's not possible to relate exon skipping event to the particular full length isoform due to the short reads length, we can still speculate on tissue specificity of the exon skipping events.

The list of differentially expressed genes related to aggressive behavior provides the basis for gene network. The usage of computer technologies such as GeneNet and ANDVisio [14] gives the possibility of reconstructing gene networks – assemblies of coordinately functioning genes controlling biochemical, molecular-genetic, and physiological processes – based on the published data. We used these tools for gene co-expression analysis and network reconstruction to reveal features of such a network based on genes related aggressive behavior.

5 Gene ontology categories for behaviour-specific genes

We used Allen Brain Atlas [6] and BioGPS databases for compilation of gene set with high expression in the brain areas studied. These databases contain the gene expression data in a wide variety of tissues and organs in human, mouse and rat. Gene list with high expression was also determined using Affymetrix U133 microarrays with probe quality filtration [19] from BioGPS (http://biogps.org). The microarray probes with high expression rate (according to the ranking of the probes of all genes, 1%) and the genes expressed in the brain regions (hypothalamus, prefrontal cortex, etc.) and not expressed in other organs (kidney, liver, etc.) have been selected from the Affymetrix U133 microarray probes represented in the BioGPS database. After the deletion of duplicated microarray probes, about 11 thousands genes (unique identifiers) have been selected. Among these genes 15% were overexpressed in at least one region of the brain. We estimated gene ontology categories for such gene lists as background for specific studies of behavior in laboratory animals. Note that data from BioGPS from microarrays only partially cover genes and brain areas studied in our RNA-seq experiments.

For the gene lists obtained we considered gene ontology (GO) categories, the gene neighbor surrounding, the context structure of regulatory regions, overlapping with the microRNAs, short non-coding transcripts in the opposite orientation, the number of exons, and the evolutionary conservation were analyzed. The functional annotation of about 1000 genes with a high expression only in the brain was conducted by the DAVID (Database for Annotation, Visualization and Integrated Discovery) software for the gene ontology analysis (http://david.abcc.ncifcrf.gov).

It is interesting to note the non-random abundance of genes in the GO categories of protein transport, phosphoproteins, and nucleotide binding complement in the gene lists. The categories of transmission of nerve impulses and neuron development were demonstrated, which is expected for the brain regions. Almost half of the genes (45.7%) from the list are related to alternative splicing featuring the reported previously plasticity of synaptic genes category in particular.

Next, we constructed differentially expressed gene lists in the brain areas studied between aggressive and tame rats and analyzed gene ontology categories. The categories of differentially expressed genes obtained from RNA-seq data again confirmed the presence of neuronal genes and signal transduction terms. There are genes known as related to aggressive behavior, such as MaoA. We continue to work on gene network reconstruction using RNA-seq experiments on additional mouse brain structures in contrast groups of laboratory animals using digital atlas of such structures [4-5; 35].

6 Analysis of alternative splicing

An additional layer of regulatory complexity in the biological processes in brain is alternative splicing and other posttranscriptional changes that expand the variety of protein isoforms produced. There are tools for the analysis of alternative splicing in RNA-Seq data [34].

Alternatively spliced exons in rat brain samples were elucidated by rMATs tool [15; 34]. The parameters were set to: -dataType = single (non paired end tags); – read Length = 49; -anchor Length=8; -junction Length = 82. Inc Form Len= 2*(j-r+1)=2*(82-49+1)=68; Skip Form Length=34.

The joint likelihood function of the statistical model in rMATs is a combination of the binomial distribution modeling the relationship of the exon inclusion reads, exon skipping reads, and the exon inclusion level in each individual replicate (1), and the normal distribution modeling the variation of the replicate exon inclusion levels within sample group (2). Thus, the joint-likelihood function is composed of two components:

1) Exon inclusion distribution is modelled by the introducing exon inclusion rate statistic ψ_I by following equation accounting for inclusion (*I)* and skipping (*S)* read numbers and corresponding effective reads length l_I and l_S:

$$\psi_I = \frac{I/l_I}{\frac{I}{l_I}+\frac{S}{l_S}}:$$

Assuming binomial distribution for exon inclusion rate we obtain:

$$\mathrm{F}(\psi) \quad = \sum_{k=0}^{I+S} \binom{I+S}{k} f(\psi)^k (1-f(\psi))^{I+S-k} \qquad (1)$$

where $f(\psi_I)$ is normalized inclusion rate ψ_I: $f(\psi_I) = \frac{l_I \psi_I}{l_I \psi_I + l_S(1-\psi_I)}$

2) The (non-paired) replicates within the each group are used to assess the exon inclusion rate variance within the group assuming normal distribution:

$$logit(\psi_I) - Normal(\mu = logit(\psi_I), \sigma^2) \quad (2)$$

The log likelihood test *LH* checks if probability of between group deviation of exon inclusion ratios is less than pre-defined threshold e: $|\psi_{I1} - \psi_{I2}| < e$ given the within group variation (2).

For the analysis of isoforms distribution in *Grin1* loci, we used Principal Components (PC) analysis based on Pearson matrix of correlation of isoforms frequencies profiles across 12 samples considered. We used XLSTAT add-on software (www.xlstat.com) to calculate the PC routine and to plot the results.

By means of rMATs we identified several thousand alternatively spliced genes within each rat brain area studied. The numbers of statistically significant differentially spliced exons between aggressive and tame species were much more modest, and majorly occurred in MT and PGM (Table 1):

Table 1: Distribution of significant (FDR<0.05) differential exon splicing events across 3 tissues in rat brain.

Brain Regions	Alternative exons significantly deviated between tame and aggressive rats*	Alternative exons total
Hypothalamus	26 (20)	1832
MT	45 (32)	1900
PGM	34 (14)	1838

*in brackets the number of neural specific exons number is presented

Table 2: DAVID GO neurospecific categories abundance elucidated in the gene list.

GO_Term	Gene count	Genes Total	P-value	FDR
GO:0043005~neuron projection	114	1242	8.8E-18	1.3E-14
GO:0030425~dendrite	62	1242	3.9E-10	5.7E-07
GO:0030424~axon	55	1242	5.5E-10	8.2E-07
GO:0016192~vesicle-med. transport	103	1337	7.9E-14	1.5E-10
GO:0045202~synapse	**109**	**1242**	**2.5E-20**	**3.7E-17**
GO:0044456~synapse part	70	1242	9.1E-12	1.4E-08
GO:0014069~postsynaptic density	31	1242	3.9E-09	5.8E-06
GO:0030054~cell junction	93	1242	4.2E-09	6.2E-06

Table 2 shows non-random gene ontology (GO) categories abundance in alternatively spliced gene lists, in particular the neuron projection and synapse, by DAVID web-server.

Note the presence of synapse GO term in the differentially spliced gene list from mouse brain. In the course of analysis of alternative splicing we selected glutamatergic receptor protein 1 (Grin1), which encodes 8 isoforms, 3 of which has significant deviation of their isoforms profile frequencies between aggressive and tame species in PGM and MT regions.

It was previously demonstrated that repeated social defeat stress enhances long-term potentiation (LTP) in NMDA receptor-mediated glutamatergic transmission in the ventral tegmental area [36]. Notably, defeated rats display enhanced learning of contextual cues paired with cocaine experience assessed using a conditioned place preference (CPP) paradigm. Enhancement of LTP in the VTA and cocaine CPP in behaving rats both require glucocorticoid receptor activation during defeat episodes. These findings suggest that enhanced glutamatergic plasticity in the VTA may contribute, at least partially, to increased addiction vulnerability following repeated stressful experiences. Another study underlines Ca2+ intrusion into cells via NMDA under leading to chemokine receptor CXCR4 mediated neurotoxicity [37]. The blockade of l-type Ca2+ channels and NMDAR both prevent CXCR4-mediated toxicity of CXCL12 in cerebrocortical neurons. Thus, Grin1 gene is relevant to the stress induced response in neurons.

The precise homeostasis of glutamate concentrations necessary for the proper neurotransmission activity may underlie the chronic stress states such as aggression and depression [38].

The 8 isoforms of Grin1 are generated by 3 alternative exons: 5, 21 and 22 [39]. We performed PC (Principal Components) decomposition based on isoforms profile frequencies in 12 samples, presented in Figure 3.

Low glutamate levels activate adaptive stress responses that include proteins that protect neurons against more severe stress. Conversely, abnormally high levels of glutamate, resulting from increased release and/or decreased removal, cause neuronal atrophy and depression. The deregulation of the glutamatergic transmission in depression could be underlined by several factors including a decreased inhibition (γ-aminobutyric acid or serotonin) or an increased excitation (primarily within the glutamatergic system). Experimental evidence shows that the activation of N-methyl-D-aspartate receptor (NMDA) and α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA) receptors (AMPAR) can exert two opposite effects on neurogenesis and neuron survival depending on the synaptic or extrasynaptic concentration.

Chronic stress, which usually underlies experimental and clinical depression, enhances glutamate release. This over activates NMDA receptors (NMDAR) and consequently impairs AMPAR activity. Various studies show that treatment with antidepressants decreases plasma glutamate levels in depressed individuals and regulates glutamate receptors by reducing NMDAR function by decreasing the expression of its subunits and by potentiating AMPAR-mediated transmission [38].

Analysing the clustering, we deduced that exon 22 alteration is not significant for the particular case. In contrast four, exon5-exon21 alteration variants are distributed in four distinct quarters of the plot [40].

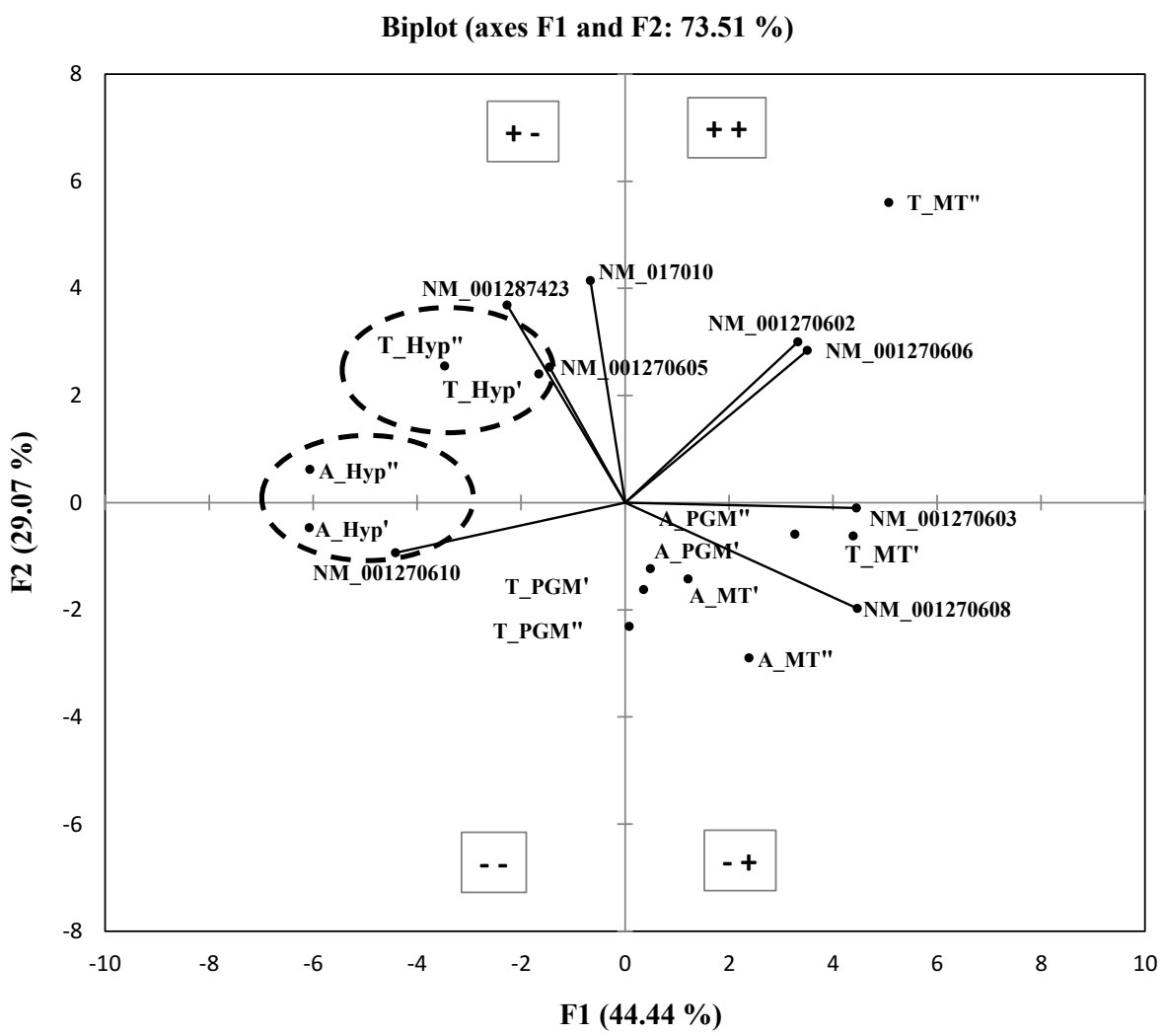

Figure 3: Isoforms expressions vectors and location of samples in Principal Components. Bracketed sign pairs (+-), (++), (-+), (--) correspond to exon 21 and 5, correspondingly; «-» denotes to absence and «+» - correspond to presence in mRNA. «A_» and «T_» prefixes designate aggressive and tame rat samples, correspondingly.

Importantly, the shortest isoform, NM_001270610, features aggressive hypothalamus brain regions, and differs statistically significantly in frequency in tame rats (FDR<8E-6). The longest isoform NM_017010, also proved to be significantly deviated (FDR<1E-2) in is characteristic of tame hypothalamus [40]. From previous studies [39] it was elucidated the exon 21 skipping mediates neuron depolarization. It could be speculated that synapse activity is higher in aggressive rats.

Discussion

Heredity has been found to significantly contribute to aggressiveness in studies of various animal species, including monkeys [41], dogs [42], mice [43], birds [44], and humans [45]. Hence, the genetic factors contribute substantially to the phenotypic variation of aggressiveness in populations. This is confirmed by inter-strain differences in the manifestation of aggression in laboratory animals and by the fast progress of selection for elevated and reduced aggressiveness traits. Note that in the scientific literature on genetics of aggressive behavior the main conclusion underlines multi-loci, or pleiotropic, determination of aggressive behavior. There are no major genes currently reported for this trait [1]. RNA profiling studies could help extend knowledge of genes related to aggressive behavior.

The genetic nature of aggressive behavior is demonstrated by differences in the manifestation of aggression between animal strains in identical environments. Inbred mouse and rat strains differ in the manifestation of aggression by the various tests: after long-term isolation; after the electric stimulation, in the tube restraint-induced attack test, and the resident – intruder test [46; 47]. These data point to the influence of genetic factors which contribute to the determination of different mechanisms controlling aggression. Studies of mice and rats revealed various relationships between the hereditary anxiety level and aggressiveness.

Currently we work on the gene network reconstruction using RNA-seq experiments on mouse brain areas in contrast groups of laboratory animals differing by behavior [35].

The functional annotation of genes with a differential expression in the brain areas of rats selected by aggressive behavior was conducted by the DAVID software for the gene ontology analysis [48] (http://david.abcc.ncifcrf.gov). Note that in scientific literature on genetics of aggressive behavior main conclusion underline multi-loci determination of aggressive behavior. There are no major genes reported. Main data sources for the laboratory animal models are Mouse Genome Database (MGD) [49] and Rat Genome Database (RGD) [50] for annotation focused on these model organisms, including knockout phenotypes and quantitative trait loci. BioGPS portal contains the data and analysis tools including arrays in tissues and cell lines [27] that could be used as background for brain gene expression studies. Regulation of gene expression in tissues could have complex patterns defined by promoter regulation, transcription factors activity, distal regulatory elements and nucleosome occupancy [51] that could be taken into account using available genomics data. For first genome-wide sequencing experiments (ChIP-seq data) skewed distributions of protein-DNA binding was shown and quantified [52]. Transcriptome profiling data face similar problems in terms of reliable detection of low expressed transcripts. Gene expression data imply development of statistical tools to find significant difference in transcript levels and gene isoform variety based on sequencing data.

Gene Atlas [6] as well as GEO NCBI resources (https://www.ncbi.nlm.nih.gov/geo/) serve as data sources for rat brain transcriptomics. New ExpGene stand-alone application helps analyze co-expression patterns. Previously we developed RatDNA database for microarray studies in rat model (age-associated deceases) [53; 54]. Based on the obtained data we constructed the MAGGENE database (Mouse and rat AGgressiveness Gene), which contains the information on genes differentially expressed in tissues of aggressive and tame rats including expression rate measured by RNA-seq data analysis. We also included the information on gene ontologies according to enrichment analysis with DAVID [48] and on gene pathways extracted from ANDVisio [31]. The SNPs from our genomic data were mapped to protein sequences. The database is available online: http://www.lcg.nsu.ru/maggene. The MySQL dump is accessible to download.

Overall, we analyzed lists of differentially expressed genes in three brain areas in samples of aggressive and tame rats selected at the Institute of Cytology and Genetics SB RAS. Computer pipelines and software for gene expression data processing helped in functional annotation of the aggressiveness-related genes. Note the presence of abundant GO categories of protein transport, phosphoproteins, and nucleotide binding in the selected list of genes. The categories of transmission of nerve impulses and neuron development were demonstrated, which is expected for the brain regions and implies deviance in synapse functioning. Almost half of the genes from the list are related to alternative splicing. RNA-seq analysis of differentially expressed genes in rat and mice confirmed the presence of genes already known as being related to aggressive behavior.

The differential alternative splicing revealed that many synapse related genes have statistically significant deviation depending on brain region and aggressive/tame status in rat. It is a novel phenomenon of the transcriptome data related to aggressive behavior [40]. We continue to work on gene network reconstruction using RNA-seq data on mouse brain structures in contrast groups of the animals different by behavior [55].

Previous studies reported quite valuable information on the gene expression alteration both in the inherited and the 'induced/acquired' models of aggression [56]. We reconfirmed the brain specific deviations of gene expression in our data, in particular in synapse specific genes. The systemic research of gene expression in the brain cells using comprehensive experimental approaches is required for interdisciplinary neurobiological studies, and future research has to be conducted using new transcriptome data and software applications for gene expression analysis [1; 16; 55]. D.A. Smagin et al [35] found differentially expressed genes in mouse behavior models of aggressive confrontations while studying larger number of brain areas. Implication of ribosomal genes in male aggressive mice neural metabolism was shown [4].

New application for network reconstruction by transcriptome data in rat hypothalamus was recently published continuing studies in the same brain areas [57]. Analysis of nucleotide polymorphisms in human genes associated with aggressive behavior by OMIM (Online Mendelian Inheritance in Man) database (http://omim.org/) revealed role of regulatory mechanisms in gene expression in promoter regions [58]. Thus, the integration of computer tools allows solving more complex problems of molecular mechanisms of behavior studies using transcriptome sequencing data.

Acknowledgements

Authors are grateful to A.L. Markel and N.N. Kudryavtseva for providing laboratory animals and behaviour experiments, to M.P. Ponomarenko, N.S. Krivosheina, K.A. Tabanykhov, E.P. Mazurina, E.S.Tyis for science discussion. The experiment was supported by the Russian Science Foundation (grant 14-14-00269) (VNB, AMS, AOB, YLO). IVC, GVO, ERG, IVM acknowledge project 0324-2015-0003 of the Russian Government Budget. Computing was supported at Shared Facility Centre "Bioinformatics", Siberian Supercomputer Center (SSCC SB RAS, www2.sscc.ru) in Novosibirsk.

References

[1] N.N. Kudryavtseva, A.L. Markel, Y.L. Orlov. Aggressive behavior: genetic and physiological mechanisms. *Vavilov journal of genetics and breeding,* 18(4/3):1133–1155, 2014. (In Russian).

[2] H.O. Heyne, S. Lautenschläger, R. Nelson, F. Besnier, M. Rotival, A. Cagan, R. Kozhemyakina, I. Plyusnina, L. Trut, Ö. Carlborg, E. Carlborg, L. Kruglyak, S. Pääbo, T. Schöneberg, F.W. Albert. Genetic influences on brain gene expression in rats selected for tameness and aggression. *Genetics*, 198(3):1277-90, 2014.

[3] N.P. Bondar, U.A. Boyarskikh, I.L. Kovalenko, M.L. Filipenko, N.N. Kudryavtseva. Molecular implications of repeated aggression: Th, Dat1, Snca and Bdnf gene expression in the VTA of victorious male mice. *PLoS ONE*, 4(1):e4190, 2009.

[4] D.A. Smagin, J.H. Park, T.V. Michurina, N. Peunova, Z. Glass, K. Sayed, N.P. Bondar, I.N. Kovalenko, N.N. Kudryavtseva, G. Enikolopov. Altered Hippocampal Neurogenesis and Amygdalar Neuronal Activity in Adult Mice with Repeated Experience of Aggression. *Front Neurosci*., 1(9):443, 2015.

[5] I.L. Kovalenko, D.A. Smagin, A.G. Galyamina, Y.L. Orlov, N.N. Kudryavtseva. [Changes in the Expression of Dopaminergic Genes in Brain Structures of Male Mice Exposed to Chronic Social Defeat Stress: An RNA-seq Study]. *MolBiol (Mosk*), 50(1):184-7, 2016. (In Russian).

[6] E.S. Lein, M.J. Hawrylycz, N. Ao et al. Genome-wide atlas of gene expression in the adult mouse brain. *Nature,* 445(7124):168-76, 2007.

[7] D.P. Calderon, M. Kilinc, A. Maritan, J.R. Banavar, D. Pfaff. Generalized CNS arousal: An elementary force within the vertebrate nervous system. *Neurosci Biobehav Rev.*, 68:167-76, 2016.

[8] D. Hitzemann, P. Bottomly, N. Darakjian, O. Walter, R. Iancu, B. Searles, S. Wilmot, McWeeney. Genes, behavior and next-generation RNA sequencing. *Genes Brain Behav*., 12(1):1-12, 2013.

[9] O.D. Iancu, A. Colville, D. Oberbeck, P. Darakjian, S.K. McWeeney R. Hitzemann. Cosplicing network analysis of mammalian brain RNA-Seq data utilizing WGCNA and Mantel correlations. *Front. Genet*., 6:174, 2015.

[10] G. Feng, P. Shaw, S.T. Rosen, S.M. Lin, W.A. Kibbe. Using the Bioconductor Gene Answers package to interpret gene lists. *Methods Mol Biol*., 802:101-112, 2012.

[11] M. Bolger, M. Lohse, B. Usadel. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15):2114-20, 2014.

[12] Trapnell, A. Roberts, L. Goff, G. Pertea, D. Kim, D.R. Kelley, H. Pimentel, S.L. Salzberg, J.L. Rinn, L. Pachter. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc*., 7(3):562-78, 2012.

[13] B. Langmead, C. Trapnell, M. Pop, S.L. Salzberg. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology*. 10(3):R25, 2009.

[14] D. Karolchik, R. Baertsch, M. Diekhans, T.S. Furey, A. Hinrichs, Y.T. Lu, K.M. Roskin, M. Schwartz, C.W. Sugnet, D.J. Thomas, R.J. Weber, D. Haussler, W.J. Kent; The UCSC genome browser database. *Nucleic acids research,* 31(1):51-54. 2003.

[15] S. Shen, J.W. Park, Z.X. Lu, L. Lin, M.D. Henry, Y.N. Wu, Q. Zhou, Y. Xing. rMATS: Robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc. Nat. Acad. Sci. USA.* 111 (51): E5593-601, 2014.

[16] A.M. Spitsina, Y.L. Orlov, N.N. Podkolodnaya, A.V. Svichkarev, A.I. Dergilev, M. Chen, N.V. Kuchin, I.G. Chernykh, B.M. Glinskij. Supercomputer analysis of genomics and transcriptomics data revealed by high-throughput DNA sequencing. *Program systems: theory and applications,* 6:1(24):157–174. 2015. (In Russian.)

[17] Y.L. Orlov, A.O. Bragin, I.V. Medvedeva, K.V. Gunbin, P.S. Demenkov, O.V. Vishnevsky, V.G. Levitsky, D.Y. Oshchepkov, N.L. Podkolodnyy, D.A. Afonnikov, I. Grosse, N.A. Kolchanov ICGenomics: Program complex for analysis of symbol genomic sequences. *Vavilov Journal of Genetics and Breeding,* 16:732–741, 2012. (In Russian).

[18] E.V. Kulakova, A.M. Spitsina, N.G. Orlova, A.I. Dergilev, A.V. Svichkarev, N.S. Safronova, I.G. Chernykh, Y.L. Orlov. Programs for analysis of genome sequencing data obtained using ChIP-seq, ChIA-PET and Hi-C technologies. *Program systems: theory and applications,* 6(2/25):129-148, 2015. (In Russian).

[19] Y.L. Orlov, J. Zhou, L. Lipovich, A. Shahab, V.A. Kuznetsov. Quality assessment of the Affymetrix U133A&B probesets by target sequence mapping and expression data analysis. *In Silico Biol*., 7(3):241-60, 2007.

[20] Y.L. Orlov, R. te Boekhorst, I.I. Abnizova. Statistical measures of the structure of genomic sequences: entropy, complexity, and position information. *J Bioinform Comput Biol*,. 4:523-36, 2006.

[21] Y.L. Orlov, V.N. Potapov. Complexity: Internet-resource for analysis of DNA sequence complexity. *Nucleic Acids Res*. 32:628-633, 2004.

[22] Y.L. Orlov, V.P. Filippov, V.N. Potapov, N.A. Kolchanov. Construction of stochastic context trees for genetic texts. *In Silico Biology*. 2(3):257-262, 2002.

[23] I. Abnizova, S. Leonard, T. Skelly, A. Brown, D. Jackson, M. Gourtovaia, G. Qi, R. te Boekhorst, N. Faruque, K. Lewis, T. Cox. Analysis of context-dependent errors for illumina sequencing. *J Bioinform Comput Biol.* 10(2):1241005, 2012.

[24] E.A. Ananko, N.L. Podkolodny, I.L. Stepanenko, O.A. Podkolodnaya, D.A. Rasskazov, D.S. Miginsky, V.A. Likhoshvai, A.V. Ratushny, N.N. Podkolodnaya, N.A. Kolchanov. GeneNet in 2005. *Nucleic Acids Res.* 33(Database issue):D425-7, 2005.

[25] E.V. Naumenko, N.K. Popova, E.M. Nikulina, N.N. Dygalo, G.T. Shishkina, P.M. Borodin, A.L. Markel. Behavior, adrenocortical activity, and brain monoamines in Norway rats selected for reduced aggressiveness towards man. *Pharmacol Biochem Behav.* 33(1):85-91, 1989.

[26] N.A. Kolchanov, M.P. Ponomarenko, A.S. Frolov, E.A. Ananko, F.A. Kolpakov, E.V. Ignatieva, O.A. Podkolodnaya, T.N. Goryachkovskaya, I.L. Stepanenko, T.I. Merkulova, V.N. Babenko, Y.V. Ponomarenko, A.V. Kochetov, N.L. Podkolodny, D.V. Vorobiev, S.V. Lavryushev, D.A. Grigorovich, Y.V. Kondrakhin, L. Milanesi, E. Wingender, V. Solovyev, G.C. Overton. 1999. Integrated databases and computer systems for studying eukaryotic gene expression. *Bioinformatics*. 15(7-8):669-86.

[27] C. Wu, X. Jin, G. Tsueng, C. Afrasiabi, A.I. Su. BioGPS: building your own mash-up of gene annotations and expression profiles. *Nucleic Acids Res*. 44(D1):D313-6, 2016.

[28] A. Franceschini, D. Szklarczyk, S. Frankild, M. Kuhn, M. Simonovic, A. Roth, J. Lin, P. Minguez, P. Bork, C. von Mering, L.J. Jensen. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res*. 41:D808-15, 2013.

[29] D. Warde-Farley, S.L. Donaldson, O. Comes, K. Zuberi, R. Badrawi, P. Chao, M. Franz, C. Grouios, F. Kazi, C.T. Lopes, A. Maitland, S. Mostafavi, J. Montojo, Q. Shao, G. Wright, G.D. Bader, Q. Morris. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res.* 38(Web Server issue):W214-20, 2010.

[30] J. Montojo, K. Zuberi, H. Rodriguez, G.D. Bader, Q. Morris. GeneMANIA: Fast gene network construction and function prediction for Cytoscape. *F1000Res*. 3:153, 2014.

[31] P.S. Demenkov, T.V. Ivanisenko, N.A. Kolchanov, V.A. Ivanisenko. ANDVisio: a new tool for graphic visualization and analysis of literature mined associative gene networks in the ANDSystem. *In Silico Biol.,* 11(3-4):149-61, 2011-2012.

[32] V.A. Ivanisenko, O.V. Saik, N.V. Ivanisenko, E.S. Tiys, T.V. Ivanisenko, P.S. Demenkov, N.A. Kolchanov. ANDSystem: an Associative Network Discovery System for automated literature mining in the field of biology. *BMC Syst Biol*., 9 Suppl 2:S2, 2015.

[33] E.Y. Bragina, E.S. Tiys, M.B. Freidin, L.A. Koneva, P.S. Demenkov, V.A. Ivanisenko, N.A. Kolchanov, V.P. Puzyrev. Insights into pathophysiology of dystrophy through the analysis of gene networks: an example of bronchial asthma and tuberculosis. *Immunogenetics*. 66(7-8):457-65, 2014.

[34] S. Shen, J.W. Park, J. Huang, K.A. Dittmar, Z.X. Lu, Q. Zhou, R.P. Carstens, Y. Xing. MATS: a Bayesian framework for flexible detection of differential alternative splicing from RNA-Seq data. *Nucleic Acids Res.,* 40(8):e61, 2012.

[35] D.A. Smagin, I.L. Kovalenko, A.G. Galyamina, A.O. Bragin, Y.L. Orlov, N.N. Kudryavtseva. Dysfunction in Ribosomal Gene Expression in the Hypothalamus and Hippocampus following Chronic Social Defeat Stress in Male Mice as Revealed by RNA-Seq. *Neural Plast.,* 2016:3289187, 2016.

[36] C.E. Stelly, M.B. Pomrenze, J.B. Cook, H. Morikawa. Repeated social defeat stress enhances glutamatergic synaptic plasticity in the VTA (ventral tegmental area) and cocaine place conditioning. *Elife*. 5. pii: e15448, 2016.

[37] A.B. Sanchez, K.E. Medders, R. Maung, P. Sánchez-Pavón, D. Ojeda-Juárez, M. Kaul. 2016. CXCL12-induced neurotoxicity critically depends on NMDA receptor-gated and L-type Ca(2+) channels upstream of p38 MAPK. *J Neuroinflammation*, 13(1):252.

[38] A. Rubio-Casillas, A. Fernández-Guasti. The dose makes the poison: from glutamate-mediated neurogenesis to neuronal atrophy and depression. *Rev Neurosci*. 27(6):599-622, 2016.

[39] J.A. Lee, Y. Xing, D. Nguyen, J. Xie, C.J. Lee, D.L. Black. Depolarization and CaM kinase IV modulate NMDA receptor splicing through two essential RNA elements. *PLoS Biol*. 5(2): e40, 2007.

[40] V.N. Babenko, A.O. Bragin, I.V. Chadaeva, Y.L. Orlov. Differential alternative splicing in brain tissues of rats selected by aggressive behavior. *Mol Biol (Mosk)*, 2017. (In press) (In Russian).

[41] L.A. Fairbanks, T.K. Newman, J.N. Bailey, M.J. Jorgensen, S.E. Breidenthal, R.A. Ophoff, A.G. Comuzzie, L.J. Martin, J. Rogers. Genetic contributions to social impulsivity and aggressiveness in vervet monkeys. *Biol. Psychiatry* 55:642-647, 2004.

[42] P. Saetre, E. Strandberg, P.E. Sundgren, U. Pettersson, E. Jazin, T.F. Bergstrom. The genetic contribution to canine personality. *Genes Brain Behav.* 5:240-248, 2006.

[43] G.A. van Oortmerssen, T.C. Bakker. Artificial selection for short and long attack latencies in wild Mus musculus domesticus. *Behav. Genet.* 11(2):115-126, 1981.

[44] P.J. Drent, K. van Oers, A.J. van Noordwijk. Realized heritability of personalities in the great tit (Parus major), *Proc. Biol. Sci.* 270:45-51, 2003.

[45] J.J. Hudziak, C.E. van Beijsterveldt, M. Bartels, M.J. Rietveld, D.C. Rettew, E.M. Derks, D.I. Boomsma. Individual differences in aggression: genetic analyses by age, gender, and informant in 3-, 7-, and 10-year-old Dutch twins. *Behav. Genet.*, 33:575-589, 2003.

[46] P.F. Brain. The adaptiveness of house mouse aggression. In: *House Mouse Aggression. A Model for Understanding the Evolution of Social Behavior*. Brain PF, Mainardi D, Parmigiani S (eds.), Harwood Acad. Publ., pp. 1-21, 1989.

[47] N.K. Popova, E.M. Nikulina, A.V. Kulikov. Genetic analysis of different kinds of aggressive behavior. *Behav Genet*. 23(5):491-7, 1993.

[48] D.W. Huang, B.T. Sherman, Q. Tan, J. Kir, D. Liu, D. Bryant, Y. Guo, R. Stephens, M.W. Baseler, H.C. Lane, R.A. Lempicki. DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists. *Nucleic Acids Res*., 35(Web Server issue):W169-75, 2007.

[49] J.A. Blake, C.J. Bult, J.T. Eppig, J.A. Kadin, J.E. Richardson, Mouse Genome Database Group. The Mouse Genome Database: integration of and access to knowledge about the laboratory mouse. *Nucleic Acids Res*., 42(Database issue):D810-7, 2014.

[50] M. Shimoyama, J. De Pons, G.T. Hayman, S.J. Laulederkind, W. Liu, R. Nigam, V. Petri, J.R. Smith, M. Tutaj, S.J. Wang, E. Worthey, M. Dwinell, H. Jacob. The Rat Genome Database 2015: genomic, phenotypic and environmental variations and disease. *Nucleic Acids Res*., 43(Database issue):D743-50, 2015.

[51] W.S. Goh, Y. Orlov, J. Li, N.D. Clarke. Blurring of high-resolution data shows that the effect of intrinsic nucleosome occupancy on transcription factor binding is mostly regional, not local. *PLoS Comput Biol*., 6(1):e1000649, 2010.

[52] V.A. Kuznetsov, Y.L. Orlov, C.L. Wei, Y. Ruan. Computational analysis and modeling of genome-scale avidity distribution of transcription factor binding sites in chip-pet experiments. *Genome Inform*., 19:83-94, 2007.

[53] O.S. Kozhevnikova, M.K. Martyschenko, M.A. Genaev, E.E. Korbolina, N.A. Muraleva, N.G. Kolosova, Y.L. Orlov. RatDNA: database on microarray studies of rats bearing genes associated with age-related diseases. *Vavilov journal of genetics and breeding*, 16(4/1):756-765, 2012. (In Russian).

[54] O.S. Kozhevnikova, E.E. Korbolina, N.A. Stefanova, N.A. Muraleva, Y.L. Orlov, N.G. Kolosova. Association of AMD-like retinopathy development with an Alzheimer's disease metabolic pathway in OXYS rats. *Biogerontology.* 14(6): 753-62, 2013.

[55] I.V. Medvedeva, O.V. Vishnevsky, N.S. Safronova, O.S. Kozhevnikova, M.A. Genaev, A.V. Kochetov, D.A. Afonnikov, Y.L. Orlov. Computer analysis of gene expression data in brain by microarrays and high-throughput sequencing. *Vavilov journal of genetics and breeding*. 17(4/1):629-638, 2013. (in Russian).

[56] B.L. Warren, V.F. Vialou, S.D. Iñiguez, L.F. Alcantara, K.N. Wright, J. Feng, P.J. Kennedy, Q. Laplant, L. Shen, E.J. Nestler, C.A. Bolaños-Guzmán. Neurobiological sequelae of witnessing stressful events in adult mice. *Biol Psychiatry,* 73(1):7-14, 2013.

[57] Y. Meng, Y. Guan, W. Zhang, Y.E. Wu, H. Jia, Y. Zhang, X. Zhang, H. Du, X. Wang. RNA-seq analysis of the hypothalamic transcriptome reveals the networks regulating physiopathological progress in the diabetic GK rat. *Sci Rep.,* 6:34138, 2016.

[58] I.V. Chadaeva, M.P. Ponomarenko, D.A. Rasskazov, E.B. Sharypova, T.V. Arshinova, P.M. Ponomarenko, L.K. Savinkova, N.A. Kolchanov. Candidate SNP markers of aggressiveness-related complications and comorbidities of genetic diseases are predicted by a significant change in the affinity of TATA-binding protein for human gene promoters. *BMC Genomics*, 2016. (In press).