Bioinformatics Analysis of Active Chromatin Allelic imbalance Across Genome: Testing Cerebral Neurons in Schizophrenia and Autism

O.I. Zolotareva[1,5], F.E. Gusev[1-3], S. Akbarian[4],

E.I. Rogaev[1-3,5]

olya.zolotareva@gmail.com

Evgeny.Rogaev@umassmed.edu

1) *Department of Human Genetics and Genomics, Vavilov Institute of General Genetics of Russian Academy of Science, Moscow, Russia*

2) *Center of Brain Neurobiology and Neurogenetics, Institute of Cytology and Genetics of Siberian Branch of Russian Academy of Sciences, Novosibirsk, Russia*

3) *Brudnick Neuropsychiatric Research Institute, Department of Psychiatry, University of Massachusetts Medical School, Worcester, Massachusetts, USA*

4) *Department of Psychiatry, Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA*

5) *Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University, Moscow 119234, Russia*

Allelic imbalance is the difference between expression levels of two alleles of a certain gene. The inherited silence of one of the two alleles in a set of mammalian genes suppressed specifically in maternal or paternal germline is termed genomic imprinting. The genomic imprinting plays a crucial role in mammalian development and is also involved in human diseases such as Prader-Willi or Silver-Russell syndromes. Apparently, allelic imbalance of gene activity in non-parent-specific manner may also occur depending on genetic and non-genetic factors. However, how commonly this phenomenon occurs for multiple genes in human populations and in different pathological and non-pathological conditions have yet to be investigated. Such study would require development of an appropriate threshold criteria and systemic approach to estimate the significance of quantitative variations among alleles using the massive genomic data of variable quality and quantity. To date, the allelic imbalance has been analyzed mainly using transcriptome data (e.g., RNA-seq data). There

are many confounding factors which may affect the correct interpretation of gene activity using RNA transcripts, including variable stability of different mRNA species and heterogeneity of the tissue material. In this study, we directly tested the allelic-based activity of gene promoters by quantitative analysis of open chromatin loci mapped across human genome. We provide (i) the bioinformatics tool to estimate the genetic-epigenetic allelic states of active-chromatin loci marked by H3K4me3 modified histones (ChIP-seq dataset); (ii) comparative analysis of allelic imbalance of promoter activity using cell-type-specific whole-genome H3K4me3 landscape of genes in prefrontal cortical neurons from multiple individuals; (iii) application of the bioinformatics tool for testing of allelic imbalance of open chromatin in neurons from autistic (ASD) [1] and schizophrenic (SZ) patients (unpublished data).

The most challenging issue was to avoid false positives in detection of allelic imbalance. Because the majority of errors rises from mapping bias effect we performed SNP-tolerated alignment. Also, before the testing, all remaining positions that are supposed to demonstrate mapping bias were excluded from further analysis. To detect allelic imbalance at low coverage and to distinguish possible allele-specific and parental-specific expression we developed a program based on Bayesian algorithm to search for patterns of imbalance between groups. Our program takes into account allele frequencies from 1000 genomes data to exclude false positives. In contrast to the majority of presented programs [e.g., 3,4] our program integrates data from multiple samples to improve statistical significance of findings located in low-covered peaks.

We performed the analysis using the ChIP-seq dataset for H3K4me3 marked loci (>20,000 across neuronal genome) generated for prefrontal cortical neural tissue cells (separated from non-neuronal cell material) from 59 individuals: 16 of them suffered from SZ, 13 were diagnosed ASD and remaining 30 were healthy controls. We revealed a set of genes demonstrating allelic imbalance of active chromatin and shifts to allele-specific and potentially parental-specific gene activity in SZ and ASD groups compared to control.

Recent studies suggested correlation between size and weight of newborns and risk to develop SZ and ASD. Above average-sized newborns have a relatively high risk to develop ASD but low risk of SZ. The reverse correlation was observed for below average-sized newborns [5]. One possible explanation is that female contribution to offspring is significantly higher than male and it is not beneficial for mother to spend all energy on a single child, therefore mother's genes try to reduce size of embryo. Potentially, correlation between SZ and ASD risks and the size of the newborn could be the result of intensification of maternal and paternal imprinting respectively. Thus, we believe that the study of allelic imbalance of gene activity in neuronal tissue could be a promising new strategy in a search for molecular factors contributing to these major psychiatric disorders.

1. H.P. Shulha, I. Cheung, C. Whittle, J. Wang, D. Virgil, C.L. Lin, Y. Guo, A. Lessard, S. Akbarian, Z. Weng (2012) Epigenetic signatures of autism: trimethylated H3K4 landscapes in prefrontal neurons, *Arch. Gen. Psychiatry,* 69(3):314-24.
2. W. Yingying, L. Xia, W. Qian-fei and J. Hongkai (2012) iASeq: integrative analysis of allele-specificity of protein-DNA interactions in multiple ChIP-seq dataset, *BMC Genomics* 2012, **13**:681.
3. R.V. Pandey, S.U. Franssen, A. Futschik, C. Schlötterer (2013) Allelic imbalance metre (Allim), a new tool for measuring allele-specific gene expression with RNA-seq data, *Mol. Ecol. Resour*, **4**:740-5.
4. J. Rozowsky, A. Abyzov, J. Wang, P Alves, D. Raha, A. Harmanci, J. Leng, R. Bjornson, Y. Kong, N. Kitabayashi, N. Bhardwaj, M. Rubin, M. Snyder, M. Gerstein. (2011). AlleleSeq: analysis of allele-specific expression and binding in a network framework. *Mol. Syst. Biol.,* **7:**522.
5. S.G. Byars, S.C. Stearns, J.J. Boomsma (2014) Opposite risk patterns for autism and schizophrenia are associated with normal variation in birth size: phenotypic support for hypothesized diametric gene-dosage effects, *Proc. Biol. Sci.,* **7:**281.