Genomics Proteomics Bioinformatics 12 (2014) 92–104

Genomics Proteomics Bioinformatics

www.elsevier.com/locate/gpb
www.sciencedirect.com

ORIGINAL RESEARCH

Mining the 3′UTR of Autism-implicated Genes for SNPs Perturbing MicroRNA Regulation

Varadharajan Vaishnavi [#], Mayakannan Manikandan [#], Arasambattu Kannan Munirajan [*]

Department of Genetics, Dr. ALM PG Institute of Basic Medical Sciences, University of Madras, Taramani Campus, Chennai 600113, India

Received 13 September 2013; revised 9 December 2013; accepted 11 January 2014
Available online 18 April 2014

Handled by Zhongming Zhao

KEYWORDS

Autism spectrum disorder;
MicroRNA;
Single nucleotide polymorphism;
3′UTR-SNP;
Gene regulation

Abstract Autism spectrum disorder (ASD) refers to a group of childhood neurodevelopmental disorders with polygenic etiology. The expression of many genes implicated in ASD is tightly regulated by various factors including microRNAs (miRNAs), a class of noncoding RNAs ~22 nucleotides in length that function to suppress translation by pairing with 'miRNA recognition elements' (MREs) present in the 3′untranslated region (3′UTR) of target mRNAs. This emphasizes the role played by miRNAs in regulating neurogenesis, brain development and differentiation and hence any perturbations in this regulatory mechanism might affect these processes as well. Recently, single nucleotide polymorphisms (SNPs) present within 3′UTRs of mRNAs have been shown to modulate existing MREs or even create new MREs. Therefore, we hypothesized that SNPs perturbing miRNA-mediated gene regulation might lead to aberrant expression of autism-implicated genes, thus resulting in disease predisposition or pathogenesis in at least a subpopulation of ASD individuals. We developed a systematic computational pipeline that integrates data from well-established databases. By following a stringent selection criterion, we identified 9 MRE-modulating SNPs and another 12 MRE-creating SNPs in the 3′UTR of autism-implicated genes. These high-confidence candidate SNPs may play roles in ASD and hence would be valuable for further functional validation.

Introduction

Autism spectrum disorder (ASD) represents a series of neurodevelopmental disorders characterized by altered social interests, communicative deficits, and restricted and repetitive behaviors, usually with an onset before the age of three [1]. Although ASD is defined by a triad of symptoms, the levels of severity and presentation may vary among individuals, demonstrating the tremendous heterogeneity of the condition.

* Corresponding author.
E-mail: akmunirajan@unom.ac.in, akmunirajan@gmail.com (Munirajan AK).
# Equal contribution.

Peer review under responsibility of Beijing Institute of Genomics, Chinese Academy of Sciences and Genetics Society of China.

ELSEVIER Production and hosting by Elsevier

http://dx.doi.org/10.1016/j.gpb.2014.01.003

The burden of ASD is increasing worldwide with recent prevalence estimates of 62 per 10,000 children [2]. Research in the last several years has substantiated the strong genetic basis and has further established the polygenic etiology of ASD [3,4]. For instance, a comprehensive transcriptomic analysis was performed to measure mRNA levels in post-mortem brains from autistic individuals and controls using Illumina microarrays. As a result, 444 genes were found to be differentially expressed [5]. This disease-specific, aberrant expression of genes may be caused by various factors, such as the presence of submicroscopic structural chromosomal alterations in the genome denoted as copy number variants (CNVs), or by perturbations in many regulatory mechanisms that govern gene expression [6–8]. Very recently, methylomic analysis of monozygotic twins discordant for ASD identified epigenetic variations that mediate disease susceptibility via altered gene dosage [9]. Besides gene dysfunction, studies reveal that the known etiological factors of ASD eventually converge toward the unifying theme of aberrant gene expression [5–9].

MicroRNAs (miRNAs) have emerged as key regulators for nearly 50% of protein-coding genes in the human genome, participating in all vital cellular processes [10–12]. miRNAs represent a class of short (~22 nt) endogenous noncoding RNAs that pairs with complementary sequences called miRNA recognition elements (MREs), which are located mostly in the 3′untranslated region (3′UTR) of target mRNAs. This miRNA:MRE pairing leads to translational inhibition or mRNA destabilization by recruiting the RNA-induced silencing complex (RISC). In mammals, complete complementarity between miRNA and MRE is rare, since a minimal 6-bp match of nucleotides 2–8 at the 5′ end of miRNA (known as the seed region) is sufficient for functionality [13]. Thus a single miRNA can target approximately 200 transcripts, and more than one miRNA can act upon a single mRNA target [14]. Importantly, almost 70% of experimentally-detectable miRNAs are expressed in the human nervous system [15]. Several miRNAs orchestrate a myriad of diverse neurodevelopmental processes including neuronal cell specification and differentiation, synaptic plasticity and memory formation [16]. Accumulating evidence suggests that alterations in miRNA expression or function are associated with the cognitive deficits and neurodevelopmental abnormalities observed in autism, schizophrenia and other forms of intellectual dysfunction [17–20]. According to a recent bioinformatics study, co-expression of neural miRNAs with their target mRNAs is a common feature [21], suggesting that any interference in miRNA:MRE interactions might affect the physiological control of gene expression. Additionally, single nucleotide polymorphisms (SNPs) in the 3′UTR of genes are recognized for their ability to perturb miRNA binding by modulating existing MREs or by creating new MREs [22–25]. The pathological significance of this unique class of functional polymorphisms is gaining recognition and their biological relevance has been explored in various diseases [26,27], including neurodegenerative diseases [28,29]. To our knowledge, MRE-SNPs have not formed the basis of any previous case-control studies in autism, and their systematic assessment might extend our understanding of the complex genetic architecture of this highly heterogeneous disorder. However, the identification of ASD-associated MRE-SNPs poses a considerable challenge due to the lack of an established model that accounts for the heterogeneity of ASD and the ever-expanding modes of miRNA-mediated gene regulation.

At this point, computational approaches are still the driving force in identifying truly functional genetic variants within MREs, and numerous databases have been developed to assist in this process. Aided by a rigorous bioinformatics approach, we identified a panel of high-confidence SNPs that either modulate existing MREs or create new MREs in the 3′UTR of autism-implicated genes. We propose that alterations in the miRNA:MRE interaction caused by these MRE-SNPs could be an alternate mechanism underlying aberrant gene expression, which may contribute to predisposition, pathogenesis, inter-individual variation in gene expression, and heterogeneity of ASD (**Figure 1**). Future experimental validations may further clarify the effect and consequences of these candidate SNPs.

Results

Study design and preliminary analyses

We followed a stepwise, integrative and stringent computational pipeline to identify and characterize MRE-SNPs in the 3′UTR of autism-implicated genes (**Figure 2**). A total of 484 genes were retrieved from the ''Human Gene Module'' of 'SFARI Gene', a comprehensive resource for obtaining reliable data on the genetics of ASD [30]. SNPs that modulate or create MREs across the 3′UTR of these genes were identified using MirSNP database [31]. The resultant list contained 13,945 SNPs in 468 out of the 484 autism-implicated genes. We chose MirSNP due to its high sensitivity in covering a majority of the validated miRNA-related SNPs when compared to other databases [31]. MirSNP also provides information on the plausible effect of SNPs on miRNA binding and categorizes them into one of the four functional classes. These include (i) break — SNP completely disrupts the binding of miRNA to the MRE, (ii) decrease — SNP reduces the binding efficacy of the miRNA to the MRE, (iii) enhance — SNP increases the binding affinity of the miRNA to the MRE and (iv) create — SNP creates new MREs for miRNAs. We grouped the first three functional classes as MRE modulating SNPs (MREm-SNPs) since the existing miRNA recognition sites were modified. The fourth category that created new MREs (MREc-SNPs) was considered as a separate group. Next, we checked the minor allele frequencies (MAFs) of these SNPs in Exome Variant Server (EVS) maintained by the National Heart, Lung and Blood Institute (NHLBI) Exome Sequencing Project. The EVS was chosen for several reasons. First, the database serves as a large repository for more than 10 million human SNPs identified by sequencing 15,336 genes in 6515 individuals of European American (EA) and African American (AA) ancestry [32]. Second, even SNPs with MAF as low as 0.1% can be identified owing to the larger sample size. Third, EVS also indicates the level of conservation and evolutionary constraints of the SNP loci by providing phylogenetic analysis with space/time model for conservation (phastCons) and genomic evolutionary rate profiling (GERP) scores. Finally, for many SNPs, the MAF is unavailable in NCBI's dbSNP. Of the 13,945 SNPs we identified, it was interesting to note that only 387 spanning the 3′UTR of 181 autism-implicated genes were documented with a MAF > 0.1% in at least one population (EA or AA). This suggests that the vast majority of MRE-SNPs are either very rare or not detected