A Single Nucleotide Polymorphism in the ARHGEF6 Gene is Associated with Increased Risk for Autism Spectrum Disorder

Matthew Pratt-Hyatt1, Rebecca A Haeusler2, Kobe C Yuen3 and William Shaw1

1The Great Plains Laboratory, Lenexa, Kansas, USA
2Department of Pathology and Cell Biology, Columbia University, New York, USA
3Stowers Institute for Medical Research, Kansas City, USA

Abstract

Background: Autism Spectrum Disorder (ASD) comprises a range of neurological conditions that can be a lifelong developmental disability. ASD is characterized by deficits in social communication and interaction with repetitive patterns of behavior and interests.

Methods: The study population consisted of 155 ASD subjects (134 males, 21 females) selected randomly from Great Plains Laboratory subjects (84 males and 21 females) as well as from the Autism Genetic Resource Exchange (50 male subjects). Subjects were diagnosed by the CARS, DSM-IV, ABC, ADI-R, ADOS, PL-ADOS, or the BSE criteria for pervasive developmental ailment and Childhood Autism Rating Scale. A total of 247 individuals were used as controls (106 males, 141 females).

Results: In our study we identified a single nucleotide polymorphism (SNP), rs2295868 in the ARHGEF6 gene (Rac/Cdc42 Guanine Nucleotide Exchange Factor 6) that was present in 36% of patients with ASD vs. 9% of controls. ARHGEF6, a Rho GTPase, is expressed mainly in the brain, immune system, and intestines.

Conclusion: SNP rs2295868 on the ARHGEF6 gene has significant association with ASD (odds ratio 4.09, p = 2.31 × 10^{-4}).

Keywords: Autism spectrum disorder; Single nucleotide polymorphism

Introduction

Autism spectrum disorder (ASD) is a neurodevelopmental disorder with a highly heritable complex biological basis. ASD is characterized by impaired social interaction, communication, and repetitive behavior [1,2]. Many studies have concluded that ASD traits are highly heritable indicating a genetic link [2-4]. Twin studies support the genetic hypothesis of autism. Monozygotic twin concordance rate has been measured to be between 60 and 90% and a dizygotic concordance rate of approximately 5% [5,6]. ASD occurs in 1:68 individuals; however there are differences between the sexes. Approximately one in 42 boys (0.0238) and one in 189 girls (0.00529) has been identified as having ASD, which gives a male to female ratio of 4.5:1 [7,8]. Multiple studies have focused on the X chromosome to help explain this discrepancy between male and female rates of ASD [9-11]. The ratio of brain development and cognition genes is higher on the X chromosome than other chromosomes [12,13]. ASD is most likely not a monogenetic disorder, instead it is most likely a neurological condition that results from multiple mutations of several genes in multiple pathways. However, finding the common risk alleles that are found in a large population of autistic individuals can help us understand the potentially vulnerable pathways in neurons in order to develop more efficient interventions for ASD.

Studies have focused on finding candidate genes that might lead to susceptibility to develop ASD. Many of these candidate genes implicated in ASD are involved in neurodevelopmental pathways, transcriptional control, and hormones [14-16]. Changes in neurite branching and dendritic spine morphology, including size, shape, and number, are hallmarks of many neurological conditions, including ASD.

This study was performed with 247 controls (106 male and 141 female) and 155 patients on the autistic spectrum (134 male and 21 female). Among controls, ethnicities were 62% Caucasian, 20% East Asian, and 19% Hispanic. Among cases, ethnicities were 57% Caucasian, 23% East Asian, and 13% Hispanic.

Keywords: Autism spectrum disorder; Single nucleotide polymorphism

Materials and Methods

Ethics statement

This study was approved by the ethics committee of WIRB. Written informed consent was obtained from all individuals and procedures had approval from institutional review board. The study protocol was performed in accordance with the Declaration of Helsinki.

Study population

The study population consisted of 155 ASD subjects (134 males, 21 females) selected randomly from Great Plains Laboratory subjects (84 males and 21 females) as well as from the Autism Genetic Resource Exchange (50 male subjects). Subjects were diagnosed by the CARS, DSM-IV, ABC, ADI-R, ADOS, PL-ADOS, or the BSE criteria for pervasive developmental ailment and Childhood Autism Rating Scale. A total of 247 individuals were used as controls (106 males, 141 females).

Genotyping

Genomic DNA was extracted from venous blood samples and saliva using the Agencourt Genfind V2 kit (Beckman Coulter, Indianapolis, IN, USA). The genomic DNA samples were captured with an inversion probe method for the SNPs in our panel (appendix A). These captured targets were sequenced on Illumina MiSeq sequencing system with 100 bp paired-end reads. Sequence results were mapped to UCSC hg19.
genomic reference using the ZiPhyr bioinformatics pipeline. All 247 individuals were genotyped for rs661426 and rs2295868. For rs616718, genotyping data was collected for 190 males (106 controls and 84 cases), and all 141 females.

**Statistical analysis 1**

All analyses were carried out using the statistical software R (Vienna Austria) [17]. SNPs were analyzed for association with ASD by logistic regression, individually and in an additive model. The logistic regression model was fitted stepwise both forward and backward. The receiver operating characteristic was analyzed to evaluate the predictive capacity of the model. To compare our minor allele frequencies to those in the 1000 Genomes project, we used dbSNP. To assess linkage disequilibrium in the 1000 Genomes project, we used LDlink [18]. To assess linkage of the minor alleles in our population, and to assess the OR for ASD in subjects discordant for rs2295868 and rs661426, we performed chi-square tests.

**Enhancer binding**

Analysis of published ChIP-seq data was described previously [19]. Briefly, ChIP-seq data of active enhancer markers H3K4me1, H3K27ac and p300, together with CTCF, RAD21, YY1 and the RNA-seq data were mapped to the UCSC hg19 human genome by Bowtie2 aligner 2.1.0, allowing uniquely mapped reads only up to two mismatches. Individual mapped data was converted to bigwig files before being uploaded to the IGV Genome Browser for visualization. Active enhancer regions were marked by the co-localization of H3K4me1, H3K27ac and p300.

**Results**

**Statistical analysis 2**

Three SNPs from the ARHgef6 gene (Rac/Cdc42 Guanine Nucleotide Exchange Factor 6) were analyzed because of their possible link to ASD. These SNPs (rs2295868, rs661426, and rs616718) were located in the ARHGEF6 gene region (Figure 1), these were subjected to statistical analysis. One other ARHGEF6 SNP rs545490 was excluded due to insufficient genotyping data. We noted that in our population, the minor allele frequencies differed from those observed in the 1000 genomes study. In our male control subjects, rs2295868 was less frequent and rs661426 was more frequent compared to 1000 genomes [20]. See Table 1 for full data.

**Logistic regression of data**

In individual analysis, two SNPs were significantly associated with ASD (Table 1): rs2295868 (p=2.31 × 10^-5, OR=4.09) and rs661426 (p=0.03, OR=1.77). We next performed stepwise fitting of a logistic regression model, using both SNPs as potential variables; both forward and backward fitting suggested that only rs2295868 predicted autism. In receiver operating characteristic analysis for rs2295868, the area under the curve (AUC) was 0.602. In females, none of the three SNPs were significantly associated with ASD (Table 1).

The loci of rs2295868 and rs661426 are 62 kilobases apart. Using the LDLink tool [18], we found that the two minor alleles are not in linkage disequilibrium in the 1000 Genomes population. In contrast, our own contingency table analysis showed tight association of the two minor alleles (p<0.0001). We analyzed the male individuals discordant for these two SNPs. Among the 88 subjects who have the major allele at rs661426, there were only 2 subjects who have the minor allele at rs2295868, and both have ASD. Among the 190 subjects who have the major allele at rs2295868, there were 95 who have the minor allele at rs661426. Of these, 50 (53.2%) were cases and 45 (46.9%) were controls (p=0.47, OR 1.29). Based on these analyses, rs661426 does not seem to contribute an independent risk of ASD, and rather, the increased risk of ASD is conferred by rs2295868.

**Enhancer binding at the ARHGEF6 locus**

We found that the rs2295868 locus was in close proximity (about 1 kb) to an active enhancer region in the intron of the ARHGEF6 gene, which was marked by the co-localization of H3K4me1, H3K27ac and p300 [21] (Figure 2). Active enhancers are sites for binding of a variety of transcription factors that are important in activating transcription. The existence of variation within the enhancer region of ARHGEF6 raises the possibility that the transcriptional status of this gene is linked to the development of ASD. However, it remains to be investigated if expression of ARHGEF6 is reduced in patients. Future experiments to study the gene expression are warranted.

**Discussion**

In this study we investigated the relationship between polymorphisms in multiple ASD risk genes. We identified one SNP rs2295868 in the ARHGEF6 gene to be associated with a risk of ASD. The OR for SNP rs2295868 in regard to autism risk is 4.09, which is higher than any other single SNP associated with autism risk [14,22-26]. Arhgef6, also known as apix or Cool-2, belongs to a family of GTPases. ARHGEF6 is expressed in multiple different tissues, but its highest levels are centered in the immune system, the brain, and the intestine [27,28]. In the brain ARHGEF6 is selectively expressed in the hippocampus CA1 region, which is important for new memory.

---

**Figure 1:** Predicted genomic structure of ARHGEF6 and domain structure of the corresponding protein. Numbered boxes indicate exons. The ATG start codon and TAA stop codon are shown. Four SNP locations (rs616718, rs661426, rs545490 and rs2295868) are indicated by arrows. The four predicted protein structures are: Calponin homology, CH; Src homology 3 domain, SH3; Dbl homology domain, DH; and pleckstrin homology domain, PH.
ARHGEF6 acts as a molecular switch for the Rho GTPases Rac1 and Cdc42 [30]. Rac1 and Cdc42 are key regulators of the actin cytoskeleton and affect diverse cellular processes, such as adhesion and migration, phagocytosis, cytokinesis, cell polarity, cell growth and survival, and neuronal morphogenesis [31,32]. ARHGEF6 is also an important promoter for the recycling of epidermal growth factor receptor (EGFR). ARHGEF6 is part of the Rho GTPase family of proteins, which are known for their role in neurite outgrowth, branching, axon pathfinding, and dendritic spine morphogenesis [33,34]. ARHGEF6 performs different roles depending on the tissue on which it is being expressed. In the brain ARHGEF6 is expressed in the hippocampal neuropil. ARHGEF6 helps to control neurological spine and neurite growth morphologies. Because of its role in neuronal growth, loss of ARHGEF6 activity results in the decrease in the density of cortical pyramidal neurons [27,30,35]. These mismanaged growth morphologies can lead to deficits in autophagy in microglia and impaired synaptic pruning, which have been linked to behaviors similar to those of ASD [36]. These deficits may contribute to the link between ARHGEF6 mutations and mental retardation [37]. One other role of ARHGEF6 induces apoptosis in cells that are experiencing oxidative stress. ARHGEF6 acts on TP53, HSPA1A, and CFLAR to include apoptosis [38]. One hypothesis for how ARHGEF6 mutations could increase risk of autism is that patients with these mutations are more chemically sensitive.

Studies have indicated that females are protected from ASD. Population screens [48-50] and high-risk sibling [48,51] studies demonstrate that there is a male-biased prevalence for ASD and that some risk factors impact males and females differently. The multiple threshold liability model of ASD, where sex is one factor, seems to indicate the existence of a “female protective effect” (FPE) [52]. The current study supports the FPE hypothesis, because of the much higher OR for autism in males with the rs2295868 allele than in females. However, because it is an X-linked gene, females with the ARHGEF6 rs2295868 SNP may be carriers for autism risk. In addition to males having a higher risk for autism with the rs2295868 allele, over half of the males with this allele who were not autistic had a diagnosis of depression and anxiety disorder.

One of the most common comorbidities of autism is gastrointestinal symptoms (GIS). Reports estimate that the frequency of GIS in the autistic community is approximately 40% [53,54]. Common symptoms are constipation, diarrhea, abdominal pain, and gastroesophageal reflux [55]. A recent study indicates that ARHGEF6 plays a role in these symptoms [28]. Chang et al. found that ARHGEF6 plays a role in autoimmune disease in the gut. They performed an X-chromosome wide association study and found that ARHGEF6 was associated with Crohn’s disease and ulcerative colitis. ARHGEF6 was also found to be highly expressed in T-cells [28]. If ARHGEF6 does play a role in GIS, it may explain why ASD and GIS are associated.

The mechanisms by which enhancers of ARHGEF6 might regulate its transcription levels are multifaceted. Some enhancers can directly interact with the gene’s promoter through enhancer-promoter looping interactions to activate transcription [56]. Studies have indicated that cross-talk between the active enhancer and promoter of ARHGEF6 will shed light on the potential mechanism of gene activation. In addition, the role of such interactions is not only important for transcription initiation but might also be essential in transcription elongation by releasing RNA polymerase II from promoter-proximal pausing [57]. Moreover, it has been reported that some enhancers are also transcribed into non-coding RNAs, known as enhancer RNAs, which could orchestrate regulation of transcription by stabilizing long-range enhancer-promoter interactions [58]. Examining how the ARHGEF6 enhancer is able to regulate transcriptional activity will allow us to decipher the molecular etiology of ASD, which might lead to potential therapeutic strategies in ameliorating the disease.
Conclusion

This study showed evidence supporting a role of the ARHGEF6 gene in the etiology of ASD. The SNP rs2295868 is located near a transcriptional enhancer site, and this polymorphism may cause a decrease in transcription of the ARHGEF6 mRNA. A decrease of ARHGEF6 mRNA may lead to multiple pathway changes because of ARHGEF6’s multifaceted role. Future studies should be done on other X-linked genes that interact with Rho GTPases such as GFD1, OPHN1, PAK3, and ARHGEF9. Other avenues of future study would be to look at patients with the rs2295868 SNP to see if patients have altered mitochondrial markers. Looking into markers such as glutathione status, mitochondrial dysfunction, or impaired energy production could produce significant results.

Acknowledgements

We gratefully acknowledge the resources provided by the Autism Genetic Resource Exchange (AGRE) Consortium and the participating AGRE families. The Autism Genetic Resource Exchange is a program of Autism Speaks and is supported, in part, by grant 1U24MH081810 from the National Institute of Mental Health to Clara M. Lajonchere (PI).

References


Table 1: Autism case-control association analysis.

<table>
<thead>
<tr>
<th>Gender</th>
<th>SNP</th>
<th>MAF in 1000 genomes</th>
<th>MAF in controls</th>
<th>MAF in cases</th>
<th>p-value</th>
<th>ORhet</th>
<th>ORhom</th>
</tr>
</thead>
<tbody>
<tr>
<td>Males</td>
<td>rs2295868</td>
<td>0.161</td>
<td>0.094</td>
<td>0.299</td>
<td>2.31 × 10⁻⁴</td>
<td>4.09</td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs661426</td>
<td>0.234</td>
<td>0.519</td>
<td>0.657</td>
<td>0.03</td>
<td>1.77</td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs16718</td>
<td>0.313</td>
<td>0.311</td>
<td>0.214</td>
<td>0.14</td>
<td>0.60</td>
<td></td>
</tr>
<tr>
<td>Females</td>
<td>rs2295868</td>
<td>0.161</td>
<td>0.220</td>
<td>0.214</td>
<td>0.99</td>
<td>1.02</td>
<td>0.84</td>
</tr>
<tr>
<td></td>
<td>rs661426</td>
<td>0.234</td>
<td>0.525</td>
<td>0.595</td>
<td>0.71</td>
<td>1.03</td>
<td>1.64</td>
</tr>
<tr>
<td></td>
<td>rs16718</td>
<td>0.313</td>
<td>0.213</td>
<td>0.167</td>
<td>0.36</td>
<td>0.57</td>
<td>1.15</td>
</tr>
</tbody>
</table>

MAF=Minor Allele Frequency, OR=Odds Ratio


