Genetic Insights from a Molecular Pathway Analysis on Two Independent Samples of Autistic Patients

Drago A*, Calabro M, Crisafulli C and Rasmussen SK

1Department of Clinical Medicine, Aarhus University, Psychiatric Forskningsenhed Vest, Herning, Denmark
2Department of Biomedical Science and Morphological and Functional Images, University of Messina, Via Consolare Valeria, 98125 Messina, Italy
3Via University College, University in Horsen, Denmark

Abstract

Background: Autism spectrum disorder (ASD) is a neurodevelopmental disorder characterized by limited interest and lacking ability in social interactions, repetitive behavior and dysfunction in social communication. ASD runs in families. Twin studies suggest a strong genetic basis for ASD. The complete definition of a genetic profile at risk for ASD is nevertheless currently lacking.

Methods: NIH-Autism datasets 3 and 4 (n=1233 and n=2890 respectively) were analyzed. A molecular pathway analysis was conducted. Quality analysis was run as usual (λ values). Plink and R (ReactomePA and Bioconductor packages) served for TDT, association tests and the molecular pathway analysis.

Results and Discussion: The “Adherens junctions’ interactions pathway” and “Axon guidance” were enriched in the first sample, while the “Extracellular matrix organization pathway” was enriched in the second sample. The “Axon guidance pathway” showed a trend for enrichment in the second sample. A trend of significant enrichment was observed for the “NCAM1 molecular pathway” when the severity of autistic symptoms was investigated.

Conclusion: Cell to cell interaction and the cell-matrix interaction may hold the genetic risk for ASD. Both neurodevelopment and immune response (T-cell) rely on those processes and may be involved in the pathophysiology of ASD.

Keywords: Autism; Genome-Wide Association Study (GWAS); Molecular pathway analysis; Enrichment

Introduction

Autism spectrum disorder (ASD) is a neurodevelopmental disorder that appears early in life and is characterized by limited interest and lacking ability in social interactions, repetitive behavior and dysfunction in social communication. Motor and intellectual deficits, together with mood and sleep disorder and sensory and gastrointestinal abnormalities are common [1]. ASD affects up to 1% of the general population [2] and has a genetic basis [3,4]. A number of genetic variations have been associated with ASD (https://gene.sfari.org/autdb/GS_Home.do), but results are not definitive [5–11]. A notable exception was recently published [12], reporting a positive association finding for a couple of genes previously associated with Autism, the SHANK3 and the WBSCR17. Apart from this result, the overall lack of replication rate in genetic association finding is in conflict with evidence showing that the heritability of ASD due to common variants is as high as 60% [13]. Limited power of the analysis, small sample sizes, and a critical phenotype definition may be limitations of the studies conducted so far [14–16]. A possible solution to this sparse genetic evidence is the “many genes common pathways” hypothesis, which suggests that the impact of different genes may converge to common pathways, whose impaired function results in the core symptoms of a disease [17]. The molecular pathways analysis is consistent with this hypothesis and may help understanding complex genetic diseases [18].

Previous results about the genetic networks whose disruption led to ASD showed that the following molecular pathways might be involved: Protein synthesis and metabolism, modulation of transcription process, chromatin remodeling, calcium signaling and the oxytocin pathway [19–21]. Nevertheless, those findings were generated by systematic reviews, and to the best of our knowledge a molecular pathway analysis on GWAS data derived from ASD trios is yet to be conducted.

In the present contribution two independent trios sample of ASD individuals are retrieved from the NIMH database (https://www.nimhgenetics.org/). A molecular pathway analysis is conducted on the transmission disequilibrium test (TDT) result in order to identify common molecular pathways enriched in variations associated with Autism in both samples. Moreover, the same approach is used to identify one or more pathways enriched in variations associated with the severity of autistic symptoms as ranked from 1 to 4 (from broad spectrum to strict Autism) in one of the samples. The present contribution is originated from a previous published abstract at the 29th ECNP Conference [22].

Materials and Methods

Dataset

Genetic data were available from the NIMH genomics (https://www.nimhgenetics.org/). The Autism Dataset 4 (Study 65/TASC GWAS Data) sample was chosen for the investigation sample. The Autism Dataset 3 (GWAS Data on 1,264 Non-AGRE Samples) served as a replication sample. GWAS were conducted with Affymetrix 5.0 in both samples.

Genetic analysis and quality control

SNPs (Single Nucleotide Polymorphism) were excluded for allele

*Corresponding author: Dr. Antonio Drago, Department of Clinical Medicine, Aarhus University, Psychiatric Forskningsenhed Vest, 7400 Herning, Denmark, Tel: +45 784 74660; E-mail: antonio.drago76@gmail.com

Received February 05, 2017; Accepted February 09 2018; Published February 13, 2018


Copyright: © 2018 Drago A, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.
frequency <0.01 and low genotype call rate. Mendelian error threshold was set at 5% for families and 10% for single variations and missing data rate was set at 5% for the Dataset. Duplicate samples, excessive mendelian errors and monzygous twin (removed one of the twins) were sample exclusion criteria. Deviations from the Hardy-Weinberg equilibrium were accepted under a P-threshold of 0.0001. Lambda values served to exclude inflation factors.

**Statistical analysis**

In the principal analysis, a transmission disequilibrium test (TDT) [23] was used to test for the presence of a genetic linkage between each and every SNP passing the quality test and the phenotype under analysis. The samples are comprised of trios were parents are labeled as "0" (cases) and the ASD child or children are labeled as "2" (controls). For the secondary analysis, an association analysis for quantitative data (command –assoc in plink) was run for every and single SNP using the phenotype from 1 to 4 available from the NIMH database, indicating different degrees of severity of autistic symptoms.

**Enrichment analysis**

Plink [24] served for the TDI GWAS analysis and genetic annotations. R [25] and dedicated packages served for the analyses. SNPs associated with the phenotype under analysis at a p level <0.05 were included in the molecular pathway analysis. The enrichment analysis was conducted in an R environment [25], through Bioconductor [26] and the package ReactomePA [27]. The Reactome [28] is a manually curated database that includes chemical reactions, biological processes and molecular pathways. Reactome PA was developed to analyze molecular pathways associations with gene lists obtained from high-throughput genomic investigations. Bonferroni, and False Discovery rate q-values are incorporated for multiple comparison corrections.

**Results**

1233 individuals were available from the Autism Dataset 3, 789 males, 444 females and 588 cases. 333 nuclear families, 3 founder singletons were detected. 579 non-founders with 2 parents in 321 nuclear families, 14 non-founders without 2 parents in 9 nuclear families and 579 affected offspring trios were detected. 393763 markers were included in the analysis. Genomic inflation was excluded (λ=1.03). The “Adherens junctions interactions pathway” was significantly enriched (p=0.000008; adj.p=0.008) (Table 1). 2890 individuals (1824 males, 1069 females) with non-missing phenotype were found in the Autism Dataset 4, 936 cases, 1954 controls and 3 missing. 1930 founders and 963 non-founders were found, the total genotyping rate was 0.93. 965 nuclear families were detected, 2 founder singletons were found and 963 non-founders with 2 parents in 963 nuclear families were included in the analysis. 934 affected offspring trios were identified. 1160305 markers were available for the analysis. Genomic inflation was excluded (λ=1.002). The “Extracellular matrix organization pathway” was significantly enriched (p=0.000008; adj.p=0.007) in this sample. The “Axon guidance pathway” showed a trend for enrichment (p=0.00019; adj.p=0.084). A trend of significant enrichment was observed for the “NCAM1 molecular pathway” when the severity of autistic symptoms was investigated (p=0.000009; adj.p=0.007) (Table 2). Finally, the “NCAM1 molecular pathway” showed a trend for association with a worse presentation of the autistic symptoms when only subjects with autism were selected out of the total sample (p=0.000009; adj.p=0.007).

**Discussion**

ASD is a frequent condition in the general population, characterized by impaired social abilities, restricted interests and repetitive behavior. The disorder was consistently proven to have a genetic basis, but the number of involved genes could be as high as hundreds, which suggests a polygenic nature. In order to test the convergence of specific molecular pathways towards ADS, a metabolic pathway analysis was undertaken in two independent samples of Autistic trios. As a result, two different molecular pathways were found to be enriched in the different databases, namely the “Adherens junctions interactions pathway” and the “Extracellular matrix organization pathway”, while a third pathway, the “NCAM1 molecular pathway” showed a trend for significance when the severity of autistic symptoms was taken into consideration. Our findings are consistent with what was anticipated to be one of the most critical molecular pathways in Autism [29]. Cell adhesion proteins are well known to influence the neuronal function. Rendall and colleagues recently reported on the activity of

<table>
<thead>
<tr>
<th>ID</th>
<th>Description</th>
<th>Gene Ratio</th>
<th>Bg Ratio</th>
<th>P-value</th>
<th>p.adjust</th>
<th>Q-value</th>
<th>Gene ID</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>381426</td>
<td>Regulation of Insulin-like Growth Factor (IGF) transport and uptake by Insulin-like Growth Factor Binding Proteins (IGFBPs)</td>
<td>7/615</td>
<td>17/6750</td>
<td>4.30E-04</td>
<td>1.50E-01</td>
<td>1.50E-01</td>
<td>F2/IGF1/IGFBP1/IGFBP3/PAPPA/PLG/PAPP2A</td>
<td>7</td>
</tr>
<tr>
<td>202430</td>
<td>Translocation of ZAP-70 to Immunological synapse</td>
<td>7/615</td>
<td>18/6750</td>
<td>6.60E-04</td>
<td>1.60E-01</td>
<td>1.60E-01</td>
<td>HLA-DQA1/HLA-DQA2/HLA-DQB1/HLA-DQB2/HLA-DRB1/HLA-DRB5</td>
<td>7</td>
</tr>
</tbody>
</table>

**Table 1: Molecular pathway analysis on Autism sample 3.**

Note: ID= Molecular pathways; ID: Description=Description of the pathway; GeneRatio=Number of genes in the pathway in the selected database/number of genes overall in the selected database; BgRatio=Number of genes in the pathway in international dataset/number of genes overall in international datasets; p.adjust=p values after Bonferroni correction; q value=p values after false discovery rate correction for multiple testing.
the CNTPAP2, a cell adhesion protein whose deletion results into myelin formation's insufficiency and delayed learning, two possible characteristics of ASD [30]. On the other hand, cell-adhesion molecules of the immunoglobulin superfamily have a critical role in brain development and in the maintenance of synaptic plasticity, and their defect can severely alter the function of the brain [31–33]. One of the most known and investigated gene located in the “Adherens junctions’ interactions pathway” is the Catenin Beta 1 (CTNNB1). Its product is a part of the proteins that constitute adherens junctions and it is involved in the Wnt pathway and the Reelin Pathway. The Wnt pathway in Autism is also gathering [34–36]. Quite interestingly, Wong and colleagues reported that the manipulation of the extracellular environment with pro-inflammatory mediators may impact both the activation of the Wnt molecular pathway and the activation of the metalloproteases that modify the extracellular matrix and allow and facilitate the cells’ migration [37]. Metalloproteases are one of the principal enzymes of the “Extracellular matrix organization pathway”. Disturbances in inflammation have been reported to be possible causes to Autism [38,39], and neurodevelopment is dependent on the efficiency of the immune system [40–42]. It is tempting to postulate, that a signal from a genetic susceptibility to an inflammation event during neurodevelopment was detected in the present contribution. A less efficient molecular cascade (the “Adherens junctions’ interactions pathway” or the “Extracellular matrix organization pathway”) would respond in a less efficient way to inflammatory insults, resulting in higher risk for ASD. This hypothesis needs to be further tested by independent analyses. Consistently with this, one of the genes that is included in the “Adherens junctions interactions pathway”, the CADM2, was previously found to be associated with ASD [43], in a sample of 1402 trios of ASD. Another gene included in the same molecular pathway, the ACTB gene, was deleted in a single case report a child exhibiting autistic like behavior along with brachycephaly, prominent ears, cryptorchidism, speech delay, poor eye contact, and outburst of aggressive behavior. Finally, NCAM1 is also a cell-adhesion molecule whose activity impacts on a wide range of events including cellular adhesion, migration, proliferation, differentiation, survival and synaptic plasticity.

Conclusion and Limitations

Adhesion molecules have been candidates for unraveling the genetics of Autism in the last decades, but results have been inconsistent or negative. A possible explanation to this apparently conflicting finding is the poor penetrance of the single variations for the phenotype under investigation. When investigated alone, SNPs do not have sufficient power to emerge from the noise signal in the most common association analyses. This caveat is addressed by a molecular pathway analysis, where the weak signals from different SNPs are gathered by the molecular pathway labeling, acquiring power to emerge as statistical significant. Molecular pathway analysis may be a promising approach to GWAS data, but it also has limits. One of the main limits of the genome wide molecular pathway analysis is that it relies on known molecular pathways and known genetic functions. Moreover, this technique is limited by the numbers of available SNPs included in the GWAS, so that the risk of false negative findings due to poor coverage of specific genes cannot be ruled out. Keeping this in mind, it is possible to combine the results of the molecular pathway analysis with the current published evidence about a specific phenotype, helping defining the genetic makeup that increases the risk for a disease or a group of diseases. Another limit of the present contribution is that it cannot take into account the de novo variations, as they may not be labeled and cannot therefore be captured in any molecular pathway. De novo variations may explain a part of the missing heritability, but it was estimated that 49% of the genetic architecture of ASD is related to common inherited variants and only 3% by de novo variations and rare variants [44].

Acknowledgements

Datasets

Genetic data were available from the NIMH Center for collaborative genetic studies.

NIMH Study 65, also known as AGP or TASC (PI: Gallagher), deposited genotype data in four sets, cleaned and raw, stage 1 and stage 2. NIMH then combined the cleaned stage1 and stage2 data into one dataset named Autism Dataset 4, in the PLINK file format. The n=2893 records in Dataset 4 occur in 1999 families (99% trios), of which n=935 families have one or more probands with a diagnosis, ranging from strict to broad autism, see documentation for definitions. The AGP Simplex Collection (TASC) was funded by an award from Autism Speaks and by funding support to the repository development by the NIMH. The principal investigator and co-investigators on this study were Louise Gallagher, Trinity College Dublin; Astrid Vicente, Instituto Gulbenkian de Ciencia, Oeiras; Joseph Buxbaum, Mount Sinai School of Medicine; Peter Szatmari, McMaster University; William McMahon, University of Utah; Michael Cuccaro, University of Miami; James Sutcliffe, Vanderbilt University; Christine Freitag, Klinikum der Johann-Wolfgang Goethe-Universität, Frankfurt/Main; Sabine Klaus, Deutsches Krebsforschungszentrum (DKFZ), Heidelberg; Veronica Vieland (DOE Director), Research Institute at Na- tionwide Children's Hospital, Ohio; Dan Geschwind, AGRE/UCLA; John Nurnberger, University of Indiana; Ed Cook, University of Illinois at Chicago; Raphael Bernier, University of Washington/CPEA. Genotype data were generated by Dan Arking and Aravinda Chakravarti using 1,264 samples that were obtained from the NIMH Center for Collaborative Genetic Studies on Mental Disorders. This study
was supported by grants from the NIH (R01 MH060007) to Aravinda Chakravarti and Dan Arking, Arking DE, Cutler DJ, Brune CW, Teslovich TM, West K, Ikeda M, Rea A, Guy M, Lin S, Cook EH, Jr., Chakravarti A, A Common Genetic Variant in the Neurexin Superfamily Member CNTNAP2 Increases Familial Risk of Autism. Am J Hum Genet 2008; 82:160-164.

Computational power
A computer cluster was used to undertake the analyses. Computational power was required for the permutation analysis. The GenomeDK HPC cluster at Aarhus University comprises of 190 nodes (3384 compute cores) connected with 10Gige/Infiniband. Each node has from 16 to 32 cores and either 64 GB, 128 GB, 256 GB, 512 GB or 1 TB of RAM. The cluster has been designed specifically for bioinformatic workloads and has a storage capacity of 3.5 PB.

Conflict of Interest
Authors declare no conflicts of interest.

References