PLOS | ONE

RESEARCH ARTICLE

A Pooled Genome-Wide Association Study of Asperger Syndrome

Varun Warrier[1‡]*, Bhismadev Chakrabarti[1,2‡], Laura Murphy[1], Allen Chan[1], Ian Craig[3], Uma Mallya[1], Silvia Lakatošová[1], Karola Rehnstrom[4], Leena Peltonen[4†], Sally Wheelwright[1], Carrie Allison[1], Simon E. Fisher[5,6], Simon Baron-Cohen[1,7]*

1 Autism Research Centre, Department of Psychiatry, University of Cambridge, Cambridge, United Kingdom, **2** School of Psychology and Clinical Language Sciences, Centre for Integrative Neuroscience and Neurodynamics, University of Reading, Reading, United Kingdom, **3** MRC Centre for Social, Genetic and Developmental Psychiatry, King's College London, Institute of Psychiatry, London, United Kingdom, **4** The Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, United Kingdom, **5** Max Planck Institute for Psycholinguistics, 6500 AH, Nijmegen, The Netherlands, **6** Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, Nijmegen, The Netherlands, **7** CLASS Clinic, Cambridgeshire and Peterborough NHS Foundation Trust (CPFT), Cambridge, United Kingdom

† Deceased.
‡ These authors are joint first authors on this work.
* vw260@medschl.cam.ac.uk (VW); sb205@cam.ac.uk (SB-C)

CrossMark
click for updates

OPEN ACCESS

Citation: Warrier V, Chakrabarti B, Murphy L, Chan A, Craig I, Mallya U, et al. (2015) A Pooled Genome-Wide Association Study of Asperger Syndrome. PLoS ONE 10(7): e0131202. doi:10.1371/journal.pone.0131202

Editor: Peristera Paschou, Democritus University of Thrace, GREECE

Received: January 29, 2015

Accepted: May 30, 2015

Published: July 15, 2015

Copyright: © 2015 Warrier 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.

Data Availability Statement: Summary data has been provided in the manuscript. Due to ethical restrictions related to patient consent, individual-level data are available upon request to Varun Warrier (vw260@medschl.cam.ac.uk).

Funding: This work was funded by grants to SB-C from the Nancy Lurie Marks Family Foundation, the Medical Research Council (MRC) UK, Target Autism Genome, and the Autism Research Trust (ART). LM and SEF were supported by the Max Planck Society. VW is funded by St. John's College, Cambridge and the Cambridge Commonwealth Trust.

Abstract

Asperger Syndrome (AS) is a neurodevelopmental condition characterized by impairments in social interaction and communication, alongside the presence of unusually repetitive, restricted interests and stereotyped behaviour. Individuals with AS have no delay in cognitive and language development. It is a subset of Autism Spectrum Conditions (ASC), which are highly heritable and has a population prevalence of approximately 1%. Few studies have investigated the genetic basis of AS. To address this gap in the literature, we performed a genome-wide pooled DNA association study to identify candidate loci in 612 individuals (294 cases and 318 controls) of Caucasian ancestry, using the Affymetrix GeneChip Human Mapping version 6.0 array. We identified 11 SNPs that had a p-value below $1x10^{-5}$. These SNPs were independently genotyped in the same sample. Three of the SNPs (rs1268055, rs7785891 and rs2782448) were nominally significant, though none remained significant after Bonferroni correction. Two of our top three SNPs (rs7785891 and rs2782448) lie in loci previously implicated in ASC. However, investigation of the three SNPs in the ASC genome-wide association dataset from the Psychiatric Genomics Consortium indicated that these three SNPs were not significantly associated with ASC. The effect sizes of the variants were modest, indicating that our study was not sufficiently powered to identify causal variants with precision.

Introduction

Asperger Syndrome (AS) is a neurodevelopmental condition and a subset of Autism Spectrum Conditions (ASC) [1]. Individuals with ASC have difficulties in social interaction and

Competing Interests: The authors have declared that no competing interests exist.

communication, alongside unusually repetitive and stereotyped behaviour and unusually narrow interests. In AS, language and cognitive development proceed on time. ASC is highly heritable [2], with monozygotic twin heritability rate estimated between 73 – 95% [3], and has a prevalence of approximately 1% [4]. ASC is characterized by high clinical and aetiological heterogeneity. Environmental, epigenetic and genetic factors have been implicated in ASC [3, 5–7]. Currently, more than 660 genes are implicated in ASC (https://gene.sfari.org/autdb/HG_Home.do), though no single gene or variant accounts for more than 1–2% of cases [5,8]. Additionally, several large copy number variants that duplicate or delete multiple genes have also been identified in association with ASC [3,9].

Though no common variant has been consistently associated with ASC across multiple Genome-wide Association Studies (GWAS) [3,5], it is clear now that they contribute considerably to the variation in ASC [10, 11]. Two recent studies have identified that common inherited variation contributes to between 40 – 60% of the variance in ASC [10, 11]. However, despite the majority of variance attributable to common inherited variants, as explained earlier, genome-wide association studies have failed to consistently identify causative variations. One explanation for this lack of success is that genome-wide association studies in ASC may be underpowered to detect small effect sizes; the largest ASC GWAS had less than 9000 participants (cases and controls) and, although this seems large, it has been argued that much larger sample sizes are needed (in the range of tens and hundreds of thousands) to successfully identify causative variants [9]. An alternative view is that the inability to consistently identify causative common variants is due to the underlying genetic and phenotypic heterogeneity. At a phenotypic level, delay and difficulties in language development is an important source of heterogeneity in ASC. Language delay in individuals with ASC is associated in changes in brain volume in both total grey matter and in specific regions in the brain [12]. This different brain architecture points to different biological and genetic networks involved in different forms of ASC. As mentioned earlier, AS is a subset of ASC where individuals have no language delay, suggesting it may have a genetic architecture distinct from the rest of ASC.

Only a few studies have specifically investigated the genetics of AS. In one of the first such studies, we tested for associations between 216 SNPs across 68 candidate genes. We identified nominal associations between SNPs in 14 genes and AS [13]. In the current study, we performed a pooled DNA genome wide association in individuals with AS and controls to identify SNPs in a hypothesis-free way. DNA pooling is a rapid, efficient and economical method to identify genetic associations in various conditions [14]. We hypothesized that genome-wide DNA pooling would detect the differences in allele frequencies between individuals with AS and controls. SNPs whose p-values were below a pre-defined threshold were then individually genotyped in the same sample using an established approach reported in several previous studies [15–17].

Methods

Participants

612 individuals were genotyped in the pooled genotyping stage. There were 294 cases (males = 254, females = 40, reflecting the male bias in AS [5]) and 318 controls (males = 250, females = 68). 607 of these individuals were individually genotyped. 5 (1 case and 4 controls) individuals were not genotyped at this stage due to poor DNA quality. All participants reported Caucasian ancestry for at least 2 generations. All cases were recruited from the Cambridge Autism Research Database (CARD) at www.autismresearchcentre.com, and reported that they had a clinical diagnosis of AS according to DSM IV or ICD-10 criteria. Clinical diagnostic assessment was done by independent clinicians. Control participants were recruited through

advertisement and reported that they were free of psychiatric and neurological conditions. Written consent was obtained from all participants. Ethical approval was obtained from the National Health Service Research Ethics Service (NRES).

Pooled DNA Genotyping

DNA from each participant was extracted from buccal swabs and anonymized. DNA was then suspended in Tris-EDTA and quantified using PicoGreen double-stranded DNA quantification reagent (Invitrogen, USA). 100 ng of DNA from each individual was added to their respective pool. The cases were divided into 7 pools with 5 pools for males and 2 pools for females. On average, there were 42 participants in each pool, though the numbers ranged from 12 to 59. Two additional pools with 24 female cases and 44 male cases were genotyped, but were not included in the analysis or taken forward for individual genotyping due to DNA contamination. The controls were divided into 9 pools with an average of 35 participants per pool. The number of participants per control pool ranged from 14 participants to 57.

Genotyping was performed using the Affymetrix GeneChip Human Mapping version 6.0 array (Affymetrix, California, USA) according to the protocol recommended by the company. Cell intensity (.cel) files were generated using GeneChip Scanner 3000 7g. The files generated were converted into relative allele signal scores (RAS) using a custom made script (snpmap.R [18]).

To test for differences in allele frequencies for each SNP between the cases and the controls, independent t-tests (equal variance assumed) were performed using the mean RAS scores from the pools. In addition, Levene's test was performed to check for equality of variance. A threshold of significance was chosen *a priori* at $p = 1 \times 10^{-5}$. This particular threshold was chosen in order to reduce the risk of false negatives due to the loss of power from DNA pooling [19*] and this is a threshold that is typically used in the discovery phase of GWAS studies. All SNPs were screened for quality control. The study design had approximately 38% power to detect variants with an effect of 1.3 for the given threshold of significance after taking power loss due to DNA pooling into consideration. The frequency of both the marker and the effect allele was 0.5 for the power calculation. SNPs were rejected if they had a minor allele frequency (MAF) below 0.01 in the Caucasian population according to the HapMap project, and if the coefficient of variation (calculated as SD/mean) in more than 50% of the pool was greater than 0.2. All SNPs which passed quality control and had a p-value below the threshold of significance were taken forward for individual genotyping to verify the result from the pooled association. Nominally significant SNPs in the individual genotyping stage were further investigated using summary genome-wide association data of the ASC cohort available from the Psychiatric Genomics Consortium (PGC, http://www.med.unc.edu/pgc/). The PGC analysed genome-wide SNPs using DNA from 161 cases, 526 controls, 4788 trio cases and 4788 trio pseudocontrols, all of Caucasian ancestry. A crucial difference in the PGC cohort from our study cohort is that the PGC cohort did not stratify for language delay (and hence includes cases of autistic disorder/childhood autism as well as AS). Additional details of methods, statistical analyses and participant ancestry are provided elsewhere [20*].

Individual Genotyping and functional annotation

Individual genotyping was performed by Geneservices UK Ltd using the Sequenom MassARRAY iPLEX platform (Sequenom, San Diego, USA). 5 (1 case and 4 controls) individuals who were genotyped at the pooled DNA analysis stage were not included in this stage due to poor DNA quality. Total genotyping rate was 97%. MAF in the genotyped sample for all the SNPs was above 0.05. Allelic association testing was performed using Plink v1.07 (http://pngu.mgh.

harvard.edu/~purcell/plink/) [21]. Functional annotation was performed using Haploreg v2 (http://www.broadinstitute.org/mammals/haploreg/) [22] and SNPnexus (http://www.snp-nexus.org/) [23].

Validation of DNA pooling and replication

To assess the accuracy of DNA pooling in predicting differences in allele frequency, we individually genotyped 12 random SNPs in all the participants. This includes 11 SNPs that did not reach the predefined threshold in the pooled association stage and one SNP, rs7785891, which did reach the threshold. Pearson's correlation coefficient between the mean RAS scores and the allele frequency was calculated at r = 0.65. This correlation is considerably higher than another study that used pooled DNA obtained from cheek swabs on the same platform, though lower than the correlation reported for DNA obtained from blood samples [24].

Results

In the DNA-pooling stage, 11 SNPS passed the threshold of significance and quality control. Additionally, 5 SNPs with p-values below 1 x 10^{-5} failed quality control at the pooling stage (Fig 1). All the 11 SNPs were individually genotyped and all the SNPs passed quality control in the individual genotyping phase. Three SNPs were nominally significant at $p < 0.05$ in this stage (rs7785891, rs1268055, and rs2782448). None of the SNPs survived correction for multiple testing using Bonferroni correction. None of these three nominally significant SNPs from the individual genotyping stage were significant in the PGC ASC dataset. Results are summarised in Table 1. A Q-Q plot of the results from the pooling stage is provided in Fig 2.

Discussion

The current study used pooled DNA analysis to identify common variants associated with AS. Using pooled DNA we scanned the genome for SNPs that had a difference in allele frequencies between the case groups and the control groups. A threshold of 1 x 10^{-5} was selected in the pooling stage due to the loss of power during DNA pooling and to control for false negatives. SNPs which had p-values below the pre-defined threshold were treated as candidate SNPs and genotyped individually in the same group of individuals to more accurately estimate allele frequencies. Of the 11 SNPs that crossed the threshold of significance in the pooling stage, only three remained nominally significant after the individual genotyping stage. These three SNPs were not significantly associated with ASC in a larger, more heterogeneous ASC cohort from the PGC consortium.

rs778589, the top performing SNP at the pooling stage, is an intronic SNP in *DOCK4*, a gene previously associated with ASC [25, 26]. rs2782448 is an intergenic SNP at 13q21. It is

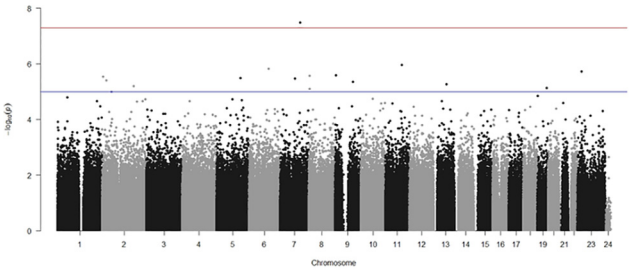

Fig 1. Manhattan plot of the SNPs tested in the pooled DNA association stage.

doi:10.1371/journal.pone.0131202.g001

Table 1. Summary of results.

RSID	Chr	Closest Gene	P-value (pooled genotyping)	P-value (individual genotyping)	Allele 1 (minor allele)	Freq of allele 1	Allele 2 (major allele)	Freq of allele 2	Chi-Sq	Odds Ratio	P-value (PGC cohort)
rs2782448	13	*KLHL1*	5.32759E-06	0.005566	G	0.2872	C	0.7128	7.686	0.7075	0.4867
rs1268055	6	*ARMC2*	1.49962E-06	0.03515	G	0.4468	A	0.5532	4.438	0.782	0.3917
rs7785891	7	*ZNF277/DOCK4*	3.27832E-08	0.04701	G	0.4094	C	0.5906	3.945	1.268	0.4532
rs25870	5	*FSTL4*	3.23677E-06	0.09446	C	0.2555	T	0.7445	2.797	0.8021	NA
rs7826102	8	*CSMD1 /LOC100287015*	2.73904E-06	0.1654	G	0.09464	C	0.90536	1.924	0.7685	NA
rs4665507	2	*LOC645949 / KLHL29*	4.00377E-06	0.5449	G	0.1098	A	0.8902	0.3665	0.8959	NA
rs7047415	9	*PTCH1 /LINC00476*	4.41161E-06	0.576	T	0.1036	C	0.8964	0.3127	1.116	NA
rs11901152	2	*NRXN1 / ASB3 /*	9.99301E-06	0.4843	A	0.2927	G	0.7073	0.4892	0.9155	NA
rs1036557	8	*CSMD1*	7.96515E-06	0.5411	G	0.4173	C	0.5827	0.3735	0.9298	NA
rs13005010	2	*ITGA6*	6.31853E-06	0.2523	A	0.113	G	0.887	1.31	1.249	NA
rs1526483	7	*SEMA3A*	3.40181E-06	0.908	G	0.1232	A	0.8768	0.01336	0.9796	NA

doi:10.1371/journal.pone.0131202.t001

371 kb from *KLHL1* and 7.5 kb from the 3' end of RP11- 459J23.1, a LincRNA identified by the Gencode project. 13q21 has been previously implicated in both autism [27, 28] and Specific Language Impairment [29]. The third nominally significant SNP, rs1268055, is an intronic SNP in *ARMC2*, a gene with uncertain function in humans.

The major limitation of this study is power. First, DNA pooling retains only 68% percent of the power [19]. Second, even after only selecting for individuals with AS, no SNP remained significant after Bonferroni correction. This indicates that larger sample sizes are required to detect causative alleles of small effect sizes. There is a considerably high correlation between the two stages of analysis, yet of the eleven SNPs selected for individual genotyping, only three remained nominally significant at this stage. The top two associated SNPs that passed quality control in the pooled DNA analysis stage were both nominally significant at the individual genotyping stage. However, they did not remain significant after correction for multiple comparisons. Of the three nominally significant alleles, rs7785891 has an odds ratio above 1, whereas rs1268055 and rs2782448 have odds ratio below 1 (see Table 1). Finally, while all our participants reported Caucasian ancestry for at least two generations, population stratification can confound the results and lead to false positives [30]. There are currently no known methods to correct for population stratification for pooled DNA association studies, taking into account the polygenicity of the condition.

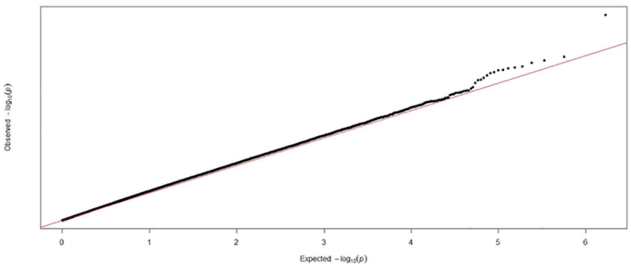

Fig 2. Quantile-quantile plot of the SNPs tested in the pooled DNA association stage.

doi:10.1371/journal.pone.0131202.g002

While the current study tested for association with AS, we also checked to see if the three nominally significant SNPs were significant in an ASC cohort. The direction of effect for all three SNPs was similar to the effect direction in our sample. However, none of the three SNPs were nominally significant in the PGC ASC cohort. This may be due to a) the heterogeneity of the PGC cohort compared to our study cohort, since the former did not stratify for language delay, and/or b) the design of the association study being different (a family based association study based on trios vs a population based study), which may lead to different signal-to-noise ratios. However, it needs to be highlighted that the effect sizes for the SNPs in both the samples were small. This underscores the need for larger sample sizes to effectively identify common variants.

In conclusion, we report the identification three SNPs (rs1268055, rs7785891 and rs2782448) as nominally associated with AS using a genome-wide pooled DNA association study. rs2782448 and rs1268055 lie in genetic loci previously implicated in ASC. None of the SNPs remained significant after Bonferroni correction, underscoring the need for larger sample sizes to uncover alleles with small effect sizes. This is the first genome-wide case-control association study to test common variants for association with AS.

Acknowledgments

We are grateful to Jon Breidbord, Lindsey Kent and Frank Dudbridge for help, advice, and discussions. Professor Peltonen kindly facilitated the collaboration with the Wellcome Trust Sanger Institute, but tragically passed away before the study was completed.

Author Contributions

Conceived and designed the experiments: BC IC LP SB-C. Performed the experiments: UM SW CA KR BC. Analyzed the data: VW AC SL LM BC. Contributed reagents/materials/analysis tools: SEF SB-C. Wrote the paper: VW BC SB-C.

References

1. American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders: DSM-IV. 4th edition. Washington, DC: American Psychiatric Association; 1994.
2. Hallmayer J, Cleveland S, Torres A, Phillips J, Cohen B, Torigoe T, et al. Genetic heritability and shared environmental factors among twin pairs with autism. Arch Gen Psychiatry. 2011; 68: 1095–102. doi: 10.1001/archgenpsychiatry.2011.76 PMID: 21727249
3. Persico AM, Napolioni V. Autism genetics. Behav Brain Res. 2013; 251:95–112. doi: 10.1016/j.bbr.2013.06.012 PMID: 23769996
4. Baron-Cohen S, Scott FJ, Allison C, Williams J, Bolton P, Matthews FE, et al. Prevalence of autism-spectrum conditions: UK school-based population study. Br J Psychiatry. 2009; 194: 500–9. doi: 10.1192/bjp.bp.108.059345 PMID: 19478287
5. Lai M-C, Lombardo MV, Baron-Cohen S. Autism. Lancet. 2013; 13: 61539–1
6. Mbadiwe T, Millis RM. Epigenetics and Autism. Autism Res Treat. 2013; 2013: 826156.
7. Samaco RC, Hogart A, LaSalle JM. Epigenetic overlap in autism-spectrum neurodevelopmental disorders: MECP2 deficiency causes reduced expression of UBE3A and GABRB3. Hum Mol Genet. 2005; 14: 483–92. PMID: 15615769
8. Bill BR, Geschwind DH. Genetic advances in autism: heterogeneity and convergence on shared pathways. Curr Opin Genet Dev. 2009; 19: 271–8. doi: 10.1016/j.gde.2009.04.004 PMID: 19477629
9. Geschwind DH. Advances in autism. Annu Rev Med. 2009; 60: 367–80. doi: 10.1146/annurev.med.60.053107.121225 PMID: 19630577
10. Klei L, Sanders SJ, Murtha MT, Hus V, Lowe JK, Willsey AJ, et al. Common genetic variants, acting additively, are a major source of risk for autism. Mol Autism. 2012; 3:9. doi: 10.1186/2040-2392-3-9 PMID: 23067556

11. Gaugler T, Klei L, Sanders SJ, Bodea CA, Goldberg AP, Lee AB, et al. Most genetic risk for autism resides with common variation. Nat Genet. 2014; 46:881–5. doi: 10.1038/ng.3039 PMID: 25038753
12. Lai MC, Lombardo MV, Ecker C, Chakrabarti B, Suckling J, Bullmore ET, et al. Neuroanatomy of Individual Differences in Language in Adult Males with Autism. Cereb Cortex. 2014. [Epub ahead of print].
13. Chakrabarti B, Dudbridge F, Kent L, Wheelwright S, Hill-Cawthorne G, Allison C, et al. Genes related to sex steroids, neural growth, and social-emotional behavior are associated with autistic traits, empathy, and Asperger syndrome. Autism Res. 2009; 2: 157–77. doi: 10.1002/aur.80 PMID: 19598235
14. Szelinger S, Pearson JV, Craig DW. Microarray-based genome-wide association studies using pooled DNA. Methods Mol Bio. 2011; 700: 49–60.
15. Butcher LM, Meaburn E, Liu L, Fernandes C, Hill L, Al-Chalabi A, et al. Genotyping pooled DNA on microarrays: a systematic genome screen of thousands of SNPs in large samples to detect QTLs for complex traits. Behav Genet. 2004; 34:549–55. PMID: 15319578
16. Butcher LM, Meaburn E, Knight J, Sham PC, Schalkwyk LC, Craig IW, et al. SNPs, microarrays and pooled DNA: identification of four loci associated with mild mental impairment in a sample of 6000 children. Hum Mol Genet. 2005; 4:1315–25.
17. Meaburn EL, Harlaar N, Craig IW, Schalkwyk LC, Plomin R. Quantitative trait locus association scan of early reading disability and ability using pooled DNA and 100K SNP microarrays in a sample of 5760 children. Mol Psychiatry. 2008; 13:729–40. PMID: 17684495
18. Davis OSP, Plomin R, Schalkwyk LC. The SNPMaP package for R: a framework for genome-wide association using DNA pooling on microarrays. Bioinformatics. 2009; 25: 281–3. doi: 10.1093/bioinformatics/btn587 PMID: 19008252
19. Barratt BJ, Payne F, Rance HE, Nutland S, Todd JA, Clayton DG. Identification of the sources of error in allele frequency estimations from pooled DNA indicates an optimal experimental design. Ann Hum Genet. 2002; 66: 393–405. PMID: 12485472
20. Cross-Disorder Group of the Psychiatric Genomics Consortium. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. Lancet. 2013; 381:1371–9 doi: 10.1016/S0140-6736(12)62129-1 PMID: 23453885
21. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007; 81: 559–75. PMID: 17701901
22. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res. 40: D930–4. doi: 10.1093/nar/gkr917 PMID: 22064851
23. Chelala C, Khan A, Lemoine NR. SNPnexus: a web database for functional annotation of newly discovered and public domain single nucleotide polymorphisms. Bioinformatics, 2008; 25: 655–661. doi: 10.1093/bioinformatics/btn653 PMID: 19098027
24. Schosser A, Pirlo K, Gaysina D, Cohen-Woods S, Schalkwyk LC, Elkin A, Korszun A, Gunasinghe C, Gray J, Jones L, Meaburn E, Farmer, AE, Craig IW, McGuffin P. Utility of the pooling approach as applied to whole genome association scans with high-density Affymetrix microarrays. BMC Res Notes. 2010; 3:274 doi: 10.1186/1756-0500-3-274 PMID: 21040578
25. Pagnamenta AT, Bacchelli E, de Jonge MV, Mirza G, Scerri TS, Minopoli F, et al. Characterization of a family with rare deletions in CNTNAP5 and DOCK4 suggests novel risk loci for autism and dyslexia. Biol Psychiatry. 2010; 4: 320–8
26. Maestrini E, Pagnamenta AT, Lamb JA, Bacchelli E, Sykes NH, Sousa I, et al. High-density SNP association study and copy number variation analysis of the AUTS1 and AUTS5 loci implicate the IMMP2L-DOCK4 gene region in autism susceptibility. Mol Psychiatry. 2010; 9: 954–68.
27. Bartlett CW, Flax JF, Logue MW, Smith BJ, Vieland VJ, Tallal P, et al. Examination of potential overlap in autism and language loci on chromosomes 2, 7, and 13 in two independent samples ascertained for specific language impairment. Hum Hered. 2004; 57: 10–20. PMID: 15133308
28. Talebizadeh Z, Arking DE, Hu VW. A Novel Stratification Method in Linkage Studies to Address Inter- and Intra-Family Heterogeneity in Autism. PloS One. 2013; 8: e67569. PMID: 23840741
29. Bartlett CW, Flax JF, Logue MW, Vieland VJ, Bassett AS, Tallal P, et al. A major susceptibility locus for specific language impairment is located on 13q21. Am J Hum Genet. 2002; 71: 45–55. PMID: 12048648
30. Tian C, Gregersen PK, Seldin MF. Accounting for ancestry: population substructure and genome-wide association studies. Hum Mol Genet. 2008; 17:R143–50. doi: 10.1093/hmg/ddn268 PMID: 18852203