npg

Open

ORIGINAL ARTICLE

Predicting the diagnosis of autism spectrum disorder using gene pathway analysis

E Skafidas[1], R Testa[2,3], D Zantomio[4], G Chana[5], IP Everall[5] and C Pantelis[2,5]

Autism spectrum disorder (ASD) depends on a clinical interview with no biomarkers to aid diagnosis. The current investigation interrogated single-nucleotide polymorphisms (SNPs) of individuals with ASD from the Autism Genetic Resource Exchange (AGRE) database. SNPs were mapped to Kyoto Encyclopedia of Genes and Genomes (KEGG)-derived pathways to identify affected cellular processes and develop a diagnostic test. This test was then applied to two independent samples from the Simons Foundation Autism Research Initiative (SFARI) and Wellcome Trust 1958 normal birth cohort (WTBC) for validation. Using AGRE SNP data from a Central European (CEU) cohort, we created a genetic diagnostic classifier consisting of 237 SNPs in 146 genes that correctly predicted ASD diagnosis in 85.6% of CEU cases. This classifier also predicted 84.3% of cases in an ethnically related Tuscan cohort; however, prediction was less accurate (56.4%) in a genetically dissimilar Han Chinese cohort (HAN). Eight SNPs in three genes (*KCNMB4*, *GNAO1*, *GRM5*) had the largest effect in the classifier with some acting as vulnerability SNPs, whereas others were protective. Prediction accuracy diminished as the number of SNPs analyzed in the model was decreased. Our diagnostic classifier correctly predicted ASD diagnosis with an accuracy of 71.7% in CEU individuals from the SFARI (ASD) and WTBC (controls) validation data sets. In conclusion, we have developed an accurate diagnostic test for a genetically homogeneous group to aid in early detection of ASD. While SNPs differ across ethnic groups, our pathway approach identified cellular processes common to ASD across ethnicities. Our results have wide implications for detection, intervention and prevention of ASD.

INTRODUCTION

Autism spectrum disorders (ASDs) are a complex group of sporadic and familial developmental disorders affecting 1 in 150 births[1] and characterized by: abnormal social interaction, impaired communication and stereotypic behaviors.[2] The etiology of ASD is poorly understood, however, a genetic basis is evidenced by the greater than 70% concordance in monozygotic twins and elevated risk in siblings compared with the population.[3–5] The search for genetic loci in ASD, including linkage and genome-wide association screens (GWAS), has identified a number of candidate genes and loci on almost every chromosome,[6–11] with multiple hotspots on several chromosomes (for example, *CNTNAP2*, *NGLNX4*, *NRXN1*, *IMMP2L*, *DOCK4*, *SEMA5A*, *SYNGAP1*, *DLGAP2*, *SHANK2* and *SHANK3*),[7,12–15] and copy number variations.[9,13,16–21] However, none of these have provided adequate specificity or accuracy to be used in ASD diagnosis. Novel approaches are required[22] to examine multiple genetic variants and their additive contribution[19,23,24] taking into account genetic differences between ethnicities and consideration of protective versus vulnerability single-nucleotide polymorphisms (SNPs).

The present study interrogated the Autism Genetics Resource Exchange (AGRE)[25] SNP data with two aims: (1) to identify groups of SNPs that populate known cellular pathways that may be pathogenic or protective for ASD, and (2) to apply machine learning to identified SNPs to generate a predictive classifier for ASD diagnosis.[26] The results were validated in two independent samples: the US Simons Foundation Autism Research Initiative

(SFARI) and UK Wellcome Trust 1958 normal birth cohort (WTBC). This novel and strategic approach assessed the contribution of various SNPs within an additive SNP-based predictive test for ASD.

MATERIALS AND METHODS

The University of Melbourne Human Research Ethics Committee approved the study (Approval Numbers 0932503.1, 0932503.2).

Subjects

(i) Index sample: subject data from 2609 probands with ASD (including Autism, Asperger's or Pervasive Developmental Disorder-not otherwise specified, but excluding RETT syndrome and Fragile X), and 4165 relatives of probands, was available from AGRE (http://www.agre.org); 1862 probands and 2587 first-degree relatives had SNP data from the Illumina 550 platform relevant to analyses (Figure 1a). Diagnosis of ASD was made by a specialist clinician and confirmed using the Autism Diagnostic Interview Revised (ADI-R[27]). Control training data was obtained from HapMap[28] instead of relatives, as the latter may possess SNPs that predispose to ASD and skew analysis (Figures 1a and b).

(ii) Independent validation samples: 737 probands with ASD (ADI-R diagnosed) derived from SFARI; 2930 control subjects from WTBC (Figure 1b).

As SNP incidence rates vary according to ancestral heritage, HapMap data (Phase 3 NCBI build 36) was utilized to allocate individuals to their closest ethnicity. Individuals of mixed ethnicity were excluded; HapMap data has 1403896 SNPs available from 11 ethnicities. Any SNPs not included in the AGRE data measured on the Illumina 550 platform were discarded, resulting in 407420 SNPs. Mitochondrial SNPs reported in AGRE,

[1]Centre for Neural Engineering, The University of Melbourne, Parkville, VIC, Australia; [2]Melbourne Neuropsychiatry Centre, Department of Psychiatry, The University of Melbourne & Melbourne Health, Parkville, VIC, Australia; [3]Department of Psychology, Monash University, Clayton, VIC, Australia; [4]Department of Haematology, Austin Health, Heidelberg, VIC, Australia; [5]Department of Psychiatry, The University of Melbourne, Parkville, Victoria, Australia. Correspondence: Professor C Pantelis, National Neuroscience Facility (NNF), Level 3, 161 Barry Street, Carlton South, VIC 3053, Australia.
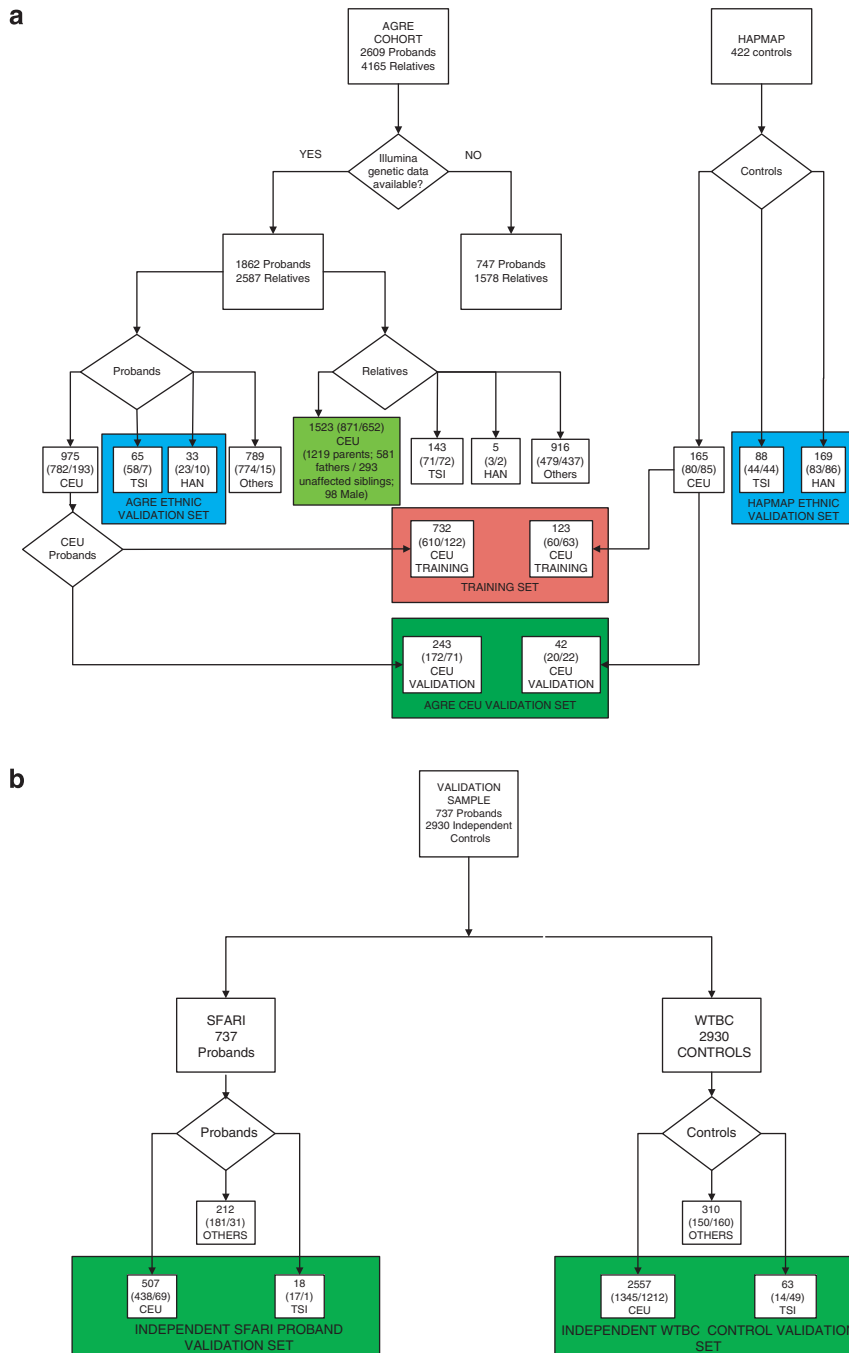E-mail: cpant@unimelb.edu.au

a

b

Figure 1. (**a** and **b**) Flow charts show the subjects used in the analyses. Key: AGRE, Autism Genetic Research Exchange; SFARI, Simons Foundation Autism Research Initiative; WTBC, Wellcome Trust 1958 normal birth cohort; CEU, of Central (Western and Northern) European origin; HAN, of Han Chinese origin; TSI, of Tuscan Italian origin; For panels 1a and b: 'red boxes'—samples used in developing the predictive algorithm; 'blue boxes'—samples used to investigate different ethnic groups; 'green boxes'—validation sets; 'light green boxes'—relatives assessed, including parents and unaffected siblings. Numbers in brackets represent numbers of males/females.

but not available in HapMap were excluded. The 30 most prevalent (>95%) SNPs within each ethnicity were identified and each ASD individual assigned to the group for which they shared the highest number of ethnically specific SNPs. HapMap groups were determined to be appropriate for analysis, as prevalence rates of the 30 SNPs relevant to each ethnicity were similar for each AGRE group assigned to that ethnicity, $P < 0.05$.

Gene set enrichment analysis (GSEA)

Pathway analysis was selected because it depicts how groups of genes may contribute to ASD etiology (Supplementary S1) and mitigates the statistical problem of conducting a large number of multiple comparisons required in GWAS studies. The current pathway analysis differs from previous ASD analyses in three unique ways: (1) we divided the cohort into ethnically homogeneous samples with similar SNP rates; (2) both protective and contributory SNPs were accounted for in the analysis and (3) the pathway test statistic was calculated using permutation analysis. Although this is computationally expensive, benefits include taking account of rare alleles, small sample sizes and familial effects. It also relaxes the Hardy–Weinberg equilibrium assumption, that allele and genotype frequencies remain constant within a population over generations. Pathways were obtained from the Kyoto Encyclopedia of Genes and Genomes (KEGG) and SNP-to-gene data obtained from the National Center

for Biotechnology Information (NCBI). Intronic and exonic SNPs were included. AGRE individuals most closely matching the genetics of Utah residents of Western and Northern European (CEU), Tuscan Italian (TSI) and Han Chinese origin were used in the analysis. CEU individuals (975 affected individuals and 165 controls) were chosen as the index sample, representing the largest group affected in AGRE (Figure 1a). The CEU and Han Chinese had 116 753 SNPs that differed, whereas the CEU and TSI had 627 SNPs, differing in allelic prevalence at $P < 1 \times 10^{-5}$. The pathway test statistic was calculated for CEU and Han individuals using a 'set-based test' in the PLINK[29] software package, with $P = 0.05$, $r^2 = 0.5$ and permutations set to at least 2 000 000. Significance threshold was set conservatively at $P < 1 \times 10^{-5}$, calculated from the number of pathways being examined (200). Therefore, significance was < 0.05/200, set at $< 1 \times 10^{-5}$ (see Supplementary S1).

Predicting ASD phenotype based upon candidate SNPs

For each individual, a 775-dimensional vector was constructed, corresponding to 775 unique SNPs identified as part of the GSEA. To examine whether SNPs could predict an individual's clinical status (ASD versus non-ASD), two-tail unpaired *t*-tests were used to identify which of the 775 SNPs had statistically significant differences in mean SNP value ($P < 0.005$). This significance level provided low classification error while maintaining acceptable variance in estimation of regression coefficients for each SNP's contribution status, and provided the set of SNPs that maximized the classifier output between the populations (Figure 2 and Supplementary S2). This resulted in 237 SNPs selected for regression analysis. Each dimension of the vector was assigned a value of 0, 1 or 3, dependent on a SNP having two copies of the dominant allele, heterozygous or two copies of the minor allele. The '0, 1, 3' weighting provided greater classification accuracy over '0, 1, 2'. Such approaches using superadditive models have been used previously to understand genetic interactions.[30] The formula for the classifier and classifier performance are presented in Supplementary S3.

The CEU sample was divided into a training set (732 ASD individuals and 123 controls) and the remainder comprised the validation set. An affected individual was given a value of 10 and an unaffected individual a value of −10, providing a sufficiently large separation to maximize the distance between means (see Supplementary S3). Least squares regression analysis of the training set determined coefficients whose sum over product by SNP value mapped SNPs to clinical status. Kolmogorov–Smirnov goodness of fit test assessed the nature of distribution of SNPs by classification. At $P = 0.05$, the distributions were accepted as being normally distributed, allowing determination of positive and negative predictive values (see ROC, Supplementary S4). The Durbin–Watson test was used to investigate the residual errors of the training set to determine if further correlations existed. At $P = 0.05$, the residuals were uncorrelated. Regression coefficients were used to assess individual SNP contribution to clinical status.

AGRE validation

After analyzing the CEU training cohort, three cohorts were used for validation: 285 (243 probands, 42 controls) CEUs; a genetically similar TSI sample (65 patients, 88 controls); and a genetically dissimilar Han Chinese population (33 patients, 169 controls). To illustrate overlap in SNPs in first-degree relatives of individuals with ASD ($n = 1512$), we mapped the SNPs of parents ($n = 1219$; 581 male) and unaffected siblings ($n = 293$; 98 male) of CEU origin who did not meet criteria for ASD. Finally, the accuracy of the predictive model was modified to test predictive ability using 10, 30 and 60 SNPs having the greatest weightings.

Independent validation

Samples included 507 CEU and 18 TSI subjects with ASD from SFARI, and 2557 CEU and 63 TSI from WTBC (Figure 1b).

RESULTS

Identification of affected pathways

Analyses focused on 975 CEU ASD individuals, in which 13 KEGG pathways were significantly affected ($P < 1 \times 10^{-5}$). The pathway analysis identified 775 significant SNPs perturbed in ASD. A number of the pathways were populated by the same genes and had inter-related functions (Table 1).
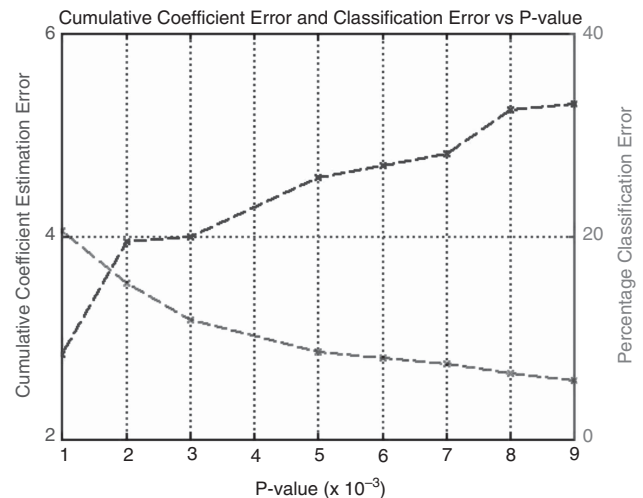

Figure 2. Cumulative coefficient estimation error and percentage classification error as a function of *P*-value; $P = 0.005$ provides good trade-off between classification performance and cumulative regression coefficient error.

The most significant pathways were: calcium signaling, gap junction, long-term depression (LTD), long-term potentiation (LTP), olfactory transduction and mitogen-activated kinase-like protein signaling. GSEA on the genetically distinct Han Chinese identified six pathways that overlapped with 13 pathways in the CEU cohort (estimate of this occurring by chance, $P = 0.05$), including: purine metabolism, calcium signaling, phosphatidylinositol signaling, gap junction, long-term potentiation and long-term depression. Related to these pathways, the statistically significant SNPs in both populations were rs3790095 within *GNAO1*, rs1869901 within *PLCB2*, rs6806529 within *ADCY5* and rs9313203 in *ADCY2*.

Diagnostic prediction of ASD

From the 775 SNPs identified within the CEU cohort, accurate genetic classification of ASD versus non-ASD was possible using 237 SNPs determined to be highly significant ($P < 0.005$). Figure 3a shows the distribution of ASD and non-ASD individuals based on genetic classification. An individual's clinical status was set to ASD if their score exceeded the threshold of 3.93. This threshold corresponds to the intersection points of the two normal curves. The theoretical classification error was 8.55%, and positive (ASD) and negative predictive values (controls) were 96.72% and 94.74%, respectively. Classification accuracy for the 285 CEU AGRE validation individuals was 85.6% and 84.3% for the TSI, while accuracy for the Han Chinese population was only 56.4%. Using the same classifier with the identical set of SNPs, accuracy of prediction of ASD in the independent data sets was 71.6%; positive and negative predictive accuracies were 70.8% and 71.8%, respectively.

SNPs were compared with the affected and unaffected individuals. Figure 3b shows that relatives (parents and unaffected siblings combined) fall between the two distributions, with a mean score of 2.68 (s.d. = 2.27). The percentage overlap of the relatives and affected individuals was 30.4%. The mean scores of the mothers and fathers did not differ (at $P = 0.05$) with scores of 2.83 (s.d. = 2.17) and 2.93 (s.d. = 2.34), respectively (see Supplementary S5), whereas unaffected siblings (not meeting diagnostic criteria for ASD) fell between parents and cases (mean = 4.74, s.d. = 3.80). In testing the robustness of the predictive model, using fewer SNPs monotonically decreased accuracy in the AGRE-CEU analyses to 72% for 60 SNPs, 58% for 30 SNPs and 53.5% for 10 SNPs, with the distribution of parents being indistinguishable from controls.

Table 1. Statistically significant pathways for the CEU and Han Chinese

KEGG pathway	Pathway name	CEU significance (P-values)	HAN significance (P-values)
hsa04020	**Calcium signaling**	**5.0×10^{-7}**	**5.0×10^{-7}**
hsa04540	**Gap junction**	**5.0×10^{-7}**	**5.0×10^{-7}**
hsa04730	**Long-term depression**	**5.0×10^{-7}**	**5.0×10^{-7}**
hsa04070	**Phosphotidylinositol signaling**	**1.5×10^{-6}**	**5.0×10^{-7}**
hsa04720	**Long-term potentiation**	**2.5×10^{-6}**	**5.0×10^{-7}**
hsa00230	**Purine metabolism**	**1.0×10^{-5}**	**5.0×10^{-7}**
hsa04010	mitogen-activated kinase-like protein	**5.0×10^{-7}**	—
hsa04740	Olfactory transduction	**5.0×10^{-7}**	—
hsa04910	Insulin signaling pathway	**1.5×10^{-6}**	—
hsa04916	Melanogenesis	**2.0×10^{-6}**	—
hsa04310	Wnt signaling	**4.0×10^{-6}**	—
hsa04912	GnRH signaling	**4.5×10^{-6}**	—
hsa04120	Ubiquitin-mediated proteolysis	**7.0×10^{-6}**	—
hsa04080	Neuroactive ligand receptor	1.2×10^{-5}	**5.0×10^{-7}**
hsa04062	Chemokine signaling pathway	1.2×10^{-5}	**5.0×10^{-7}**
hsa04060	Cytokine–cytokine receptor	1.65×10^{-5}	**5.0×10^{-7}**
hsa04114	Oocyte meiosis	—	**5.0×10^{-7}**
hsa04360	Axon guidance	—	**5.0×10^{-7}**
hsa04510	Focal adhesion	—	**5.0×10^{-7}**
hsa04514	Cell adhesion molecules	—	**5.0×10^{-7}**
hsa04670	Leukocyte transendothelial migration	—	**5.0×10^{-7}**
hsa04144	Endocytosis	—	**2.0×10^{-6}**
hsa04742	Taste transduction	—	**2.0×10^{-6}**

Abbreviations: CEU, of Central (Western and Northern) European origin; HAN, of Han Chinese origin; KEGG, Kyoto Encyclopedia of Genes and Genomes (ftp.kegg.jp).
P-values in bold are statistically significant. The pathways highlighted in 'bold' denote pathways that have reached statistical significance in both populations.
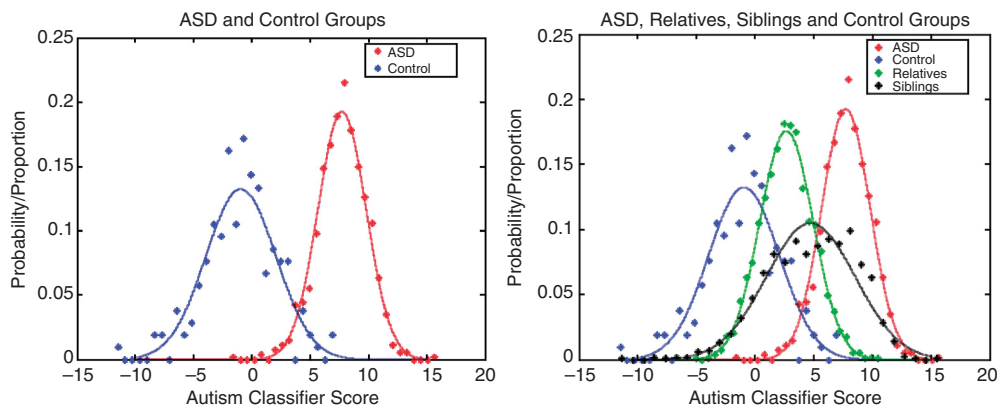

Figure 3. (**a**) Genetic-based classification of CEU population (AGRE and Controls) for ASD and non-ASD individuals, showing Gaussian approximation of distribution of individuals. As both the mapped ASD and control populations were well approximated by normal distributions, the asymptotic Test Positive Predictive Value (PPV) and Negative Predictive Value (NPV) was determined. For individuals with CEU ancestry, the PPV and NPV were 96.72% and 94.74%, respectively. (Note the test was substantially less predictive on individuals with different ancestry, that is, Han Chinese). (**b**) Genetic-based classification of CEU population, including first-degree relatives (parents and siblings of ASD children). Note that the distribution of relatives of ASD children maps between the ASD and the control groups, with no difference found between mothers and fathers (see Supplementary material S5). Key: ASD, autism spectrum disorder; relatives, first-degree relatives (parents and siblings); Siblings, siblings of ASD cases not meeting criteria for ASD; Autism Classifier Score, scores for each individual derived from the predictive algorithm, with greater values representing greater risk for autism.

Of the 237 SNPs within our classifier, presence of some contributed to vulnerability to ASD (Table 2a), whereas others were protective (Table 2b). Eight SNPs in three genes, *GRM5*, *GNAO1* and *KCNMB4*, were highly discriminatory in determining an individual's classification as ASD or non-ASD. For *KCNMB4*, rs968122 highly contributed to a clinical diagnosis of ASD, whereas rs12317962 was protective; for *GNAO1*, SNP rs876619 contributed, whereas rs8053370 was protective; for *GRM5*, SNPs rs11020772 was contributory, whereas rs905646 and rs6483362 were protective.

DISCUSSION

Using pathway analysis, we have generated a genetic diagnostic classifier based on a linear function of 237 SNPs that accurately distinguished ASD from controls within a CEU cohort. This same diagnostic classifier was able to correctly predict and identify ASD individuals with accuracy exceeding 85.6% and 84.3% in the unseen CEU and TSI cohorts, respectively. Our classifier was then able to predict ASD group membership in subjects derived from two independent data sets with an accuracy of 71.6%, thus greatly adding strength to our original finding. However, the classifier was sub-optimal at predicting ASD in the genetically distinct Han

Table 2. List of 15 most contributory (Table 2a) and 15 most protective (Table 2b) SNPs for ASD diagnosis in the CEU Cohort

SNP	Weight lower (0.95)	Weight	Weight higher (0.95)	delta	Gene number	Gene symbol
(a) Risk SNPs and their weightings						
rs968122	1.5465	1.5555	1.5645	0.0090	27 345	*KCNMB4*
rs876619	0.9476	1.2092	1.4708	0.2616	2775	*GNAO1*
rs11020772	0.8553	0.8641	0.8729	0.0088	2915	*GRM5*
rs9288685	0.5856	0.5998	0.6140	0.0142	3635	*INPP5D*
rs10193128	0.5836	0.5946	0.6056	0.0110	3635	*INPP5D*
rs7842798	0.5298	0.5386	0.5474	0.0088	114	*ADCY8*
rs3773540	0.5125	0.5208	0.5291	0.0083	55 799	*CACNA2D3*
rs1818106	0.5002	0.5161	0.5320	0.0159	80 310	*PDGFD*
rs2384061	0.4195	0.4306	0.4417	0.0111	109	*ADCY3*
rs12582971	0.3983	0.4295	0.4607	0.0312	5288	*PIK3C2G*
rs10409541	0.4067	0.4189	0.4311	0.0122	773	*CACNA1A*
rs2300497	0.3782	0.3889	0.3996	0.0107	801	*CALM1*
rs7562445	0.3741	0.3843	0.3945	0.0102	2066	*ERBB4*
rs7313997	0.3382	0.3567	0.3752	0.0185	5801	*PTPRR*
rs2239118	0.3348	0.3552	0.3756	0.0204	775	*CACNA1C*
(b) Protective SNPs and their weightings						
rs17629494	− 0.5242	− 0.5070	− 0.4898	0.0172	5592	*PRKG1*
rs4648135	− 0.5807	− 0.5260	− 0.4713	0.0547	4790	*NFKB1*
rs17643974	− 0.5527	− 0.5424	− 0.5321	0.0103	1488	*CTBP2*
rs1243679	− 0.5771	− 0.5674	− 0.5577	0.0097	341 799	*OR6S1*
rs2240228	− 0.5942	− 0.5816	− 0.5690	0.0126	26 532	*OR10H3*
rs260808	− 0.5938	− 0.5836	− 0.5734	0.0102	80 310	*PDGFD*
rs4128941	− 0.6166	− 0.6082	− 0.5998	0.0084	8313	*AXIN2*
rs769052	− 0.6321	− 0.6235	− 0.6149	0.0086	7322	*UBE2D2*
rs984371	− 0.7273	− 0.7181	− 0.7089	0.0092	219 437	*OR5L1*
rs4308342	− 1.0196	− 0.8938	− 0.7680	0.1258	1633	*DCK*
rs11145506	− 0.9400	− 0.9172	− 0.8944	0.0228	9630	*GNA14*
rs905646	− 0.9700	− 0.9624	− 0.9548	0.0076	2915	*GRM5*
rs6483362	− 0.9894	− 0.9661	− 0.9428	0.0233	2915	*GRM5*
rs12317962	− 1.4869	− 1.3200	− 1.1531	0.1669	27 345	*KCNMB4*
rs8053370	− 1.7162	− 1.6956	− 1.6750	0.0206	2775	*GNAO1*

Abbreviations: ASD, Autism spectrum disorder; CEU, of Central (Western and Northern) European origin; SNP, single-nucleotide polymorphism. Weight indicates the contribution of each SNP to ASD clinical status. 'Weight lower' indicates the 0.95 lower error bar of the estimate; 'Weight higher' indicates the 0.95 upper error bar for that SNP. Note that some genes have SNPs that contribute to risk for ASD and SNPs that protect against ASD.

Chinese cohort, which may be explained by differences in allelic prevalence. Although only 627 SNPs significantly differed between the TSI and CEU cohorts, this figure increased to 116 753 SNPs between the CEU and Han Chinese. It is likely that an additional set of SNPs may be predictive of ASD diagnosis in Han Chinese and that methods used for our classifier could be applicable to other ethnicities. Interestingly, parents and siblings of ASD-CEU individuals fell as distinct groups between the ASD and controls, reinforcing a genetic basis for ASD with neurobehavioral abnormalities reported in parents of ASD individuals also supporting our findings.[31] When we altered the classifier by reducing the number of SNPs, not only did the predictive accuracy suffer but also the relatives merged into the control group. This suggests that use of relatives as controls in SNP GWAS studies is only valid when examining small numbers of SNPs and may not be appropriate when assessing genetic interactions.

There was considerable overlap in the pathways implicated in both the CEU and Han Chinese populations. The analysis demonstrated that SNPs in the Wnt signaling pathway contributed to a diagnosis of ASD in the CEU cohort, but not in the Han Chinese population. Although of interest, a firm conclusion regarding these differences and similarities will require replication in a larger Han Chinese population. Completion of diagnostic classification studies for other ethnic groups will invariably aid in identification of common pathological mechanisms for ASD.

The SNPs contributing most to diagnosis in our classifier corresponded to genes for *KCNMB4*, *GNAO1*, *GRM5*, *INPP5D* and *ADCY8*. The three SNPs that markedly skewed an individual towards ASD were related to the genes coding for *KCNMB4*,

GNAO1 and *GRM5*. Homozygosity for *KCNMB4* SNP carries a higher risk of ASD than SNPs related to *GNAO1* and *GRM5*. By contrast, a number of SNPs protected against ASD, including rs8053370 (*GNAO1*), rs12317962 (*KCNMB4*), rs6483362 and rs905646 (*GRM5*). *KCNMB4* is a potassium channel that is important in neuronal excitability and has been implicated in epilepsy and dyskinesia.[32,33] It is highly expressed within the fusiform gyrus, as well as in superior temporal, cingulate and orbitofrontal regions (Allen Human Brain Atlas, http://human.brain-map.org/), which are areas implicated in face identification and emotion face processing deficits seen in ASD.[34] *GNAO1* protein is a subgroup of Ga(o), a G-protein that couples with many neurotransmitter receptors. Ga(o) knockout mice exhibit 'autism-like' features, including impaired social interaction, poor motor skills, anxiety and stereotypic turning behavior.[35] *GNAO1* has also been shown to have a role in nervous development co-localizing with GRIN1 at neuronal dendrites and synapses,[36] and interacting with GAP-43 at neuronal growth cones,[37] with increased levels of GAP-43 demonstrated in the white matter adjacent to the anterior cingulate cortex in brains from ASD patients.[38]

In our findings, *GRM5* SNPs have both a contributory (rs11020772) and protective (rs905646, rs6483362) effect on ASD. *GRM5* is highly expressed in hippocampus, inferior temporal gyrus, inferior frontal gyrus and putamen (Allen Human Brain Atlas), regions implicated in ASD brain MRI studies.[39] *GRM5* has a role in synaptic plasticity, modulation of synaptic excitation, innate immune function and microglial activation.[40–43] *GRM5*-positive allosteric modulators can reverse the negative behavioral effects of NMDA receptor antagonists, including stereotypies, sensory

motor gating deficits and deficits in working, spatial and recognition memory,[44] features described in ASD.[45,46] With regard to GRM5's involvement with neuroimmune function, this receptor is expressed on microglia,[40,47] with microglial activation demonstrated by us and others in frontal cortex in ASD.[48,49]

Further, as GRM5 signaling is mediated via signaling through Gene Protein Couple Receptors, a possible interaction between GNAO1 and GRM5 is plausible. Genes such as PLCB2, ADCY2, ADCY5 and ADCY8 encode for proteins involved in G-protein signaling. Given this association, GRM5 may represent a pivotal etiological target for ASD; however, further work is needed in demonstrating these potential interactions and contribution to glutamatergic dysregulation in ASD.

In conclusion, within genetically homogeneous populations, our predictive genetic classifier obtained a high level of diagnostic accuracy. This demonstrates that genetic biomarkers can correctly classify ASD from non-ASD individuals. Further, our approach of identifying groups of SNPs that populate known KEGG pathways has identified potential cellular processes that are perturbed in ASD, which are common across ethnic groups. Finally, we identified a small number of genes with various SNPs of influential weighting that strongly determined whether a subject fell within the control or ASD group. Overall these findings indicate that a SNP-based test may allow for early identification of ASD. Further studies to validate the specificity and sensitivity of this model within other ethnic groups are required. A predictive classifier as described here may provide a tool for screening at birth or during infancy to provide an index of 'at-risk status', including probability estimates of ASD-likelihood. Identifying clinical and brain-based developmental trajectories within such a group would provide the opportunity to investigate potential psychological, social and/or pharmacological interventions to prevent or ameliorate the disorder. A similar approach has been adopted in psychosis research, which has improved our understanding of the disorder and prognosis for affected individuals.[50]

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

1 Autism and Developmental Disabilities Monitoring Network Surveillance Year 2002 Principal Investigators. Prevalence of autism spectrum disorders—autism and developmental disabilities monitoring network, 14 sites, United States, 2002. MMWR Surveill Summ 2007; 56: 12–28.

2 Association AP. Diagnostic and Statistical Manual of Mental Disorders. Revised 4th edn. Washington, DC, 2000.

3 Bailey A, Le Couteur A, Gottesman I, Bolton P, Simonoff E, Yuzda E et al. Autism as a strongly genetic disorder: evidence from a British twin study. Psychol Med 1995; 25: 63–77.

4 Zhao X, Leotta A, Kustanovich V, Lajonchere C, Geschwind DH, Law K et al. A unified genetic theory for sporadic and inherited autism. Proc Natl Acad Sci USA 2007; 104: 12831–12836.

5 Cichon S, Craddock N, Daly M, Faraone SV, Gejman PV, Kelsoe J et al. Genomewide association studies: history, rationale, and prospects for psychiatric disorders. Am J Psychiatry 2009; 166: 540–556.

6 Alarcon M, Abrahams BS, Stone JL, Duvall JA, Perederiy JV, Bomar JM et al. Linkage, association, and gene-expression analyses identify CNTNAP2 as an autism-susceptibility gene. Am J Hum Genet 2008; 82: 150–159.

7 Weiss LA, Arking DE, Daly MJ, Chakravarti A. A genome-wide linkage and association scan reveals novel loci for autism. Nature 2009; 461: 802–808.

8 Sykes NH, Toma C, Wilson N, Volpi EV, Sousa I, Pagnamenta AT et al. Copy number variation and association analysis of SHANK3 as a candidate gene for autism in the IMGSAC collection. Eur J Hum Genet 2009; 17: 1347–1353.

9 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; 15: 954–968.

10 Pinto D, Pagnamenta AT, Klei L, Anney R, Merico D, Regan R et al. Functional impact of global rare copy number variation in autism spectrum disorders. Nature 2010; 466: 368–372.

11 State MW. Another piece of the autism puzzle. Nat Genet 2010; 42: 478–479.

12 Klauck SM. Genetics of autism spectrum disorder. Eur J Hum Genet 2006; 14: 714–720.

13 Szatmari P, Paterson AD, Zwaigenbaum L, Roberts W, Brian J, Liu XQ et al. Mapping autism risk loci using genetic linkage and chromosomal rearrangements. Nat Genet 2007; 39: 319–328.

14 Weiss LA, Shen Y, Korn JM, Arking DE, Miller DT, Fossdal R et al. Association between microdeletion and microduplication at 16p11.2 and autism. N Engl J Med 2008; 358: 667–675.

15 Sousa I, Clark TG, Toma C, Kobayashi K, Choma M, Holt R et al. MET and autism susceptibility: family and case-control studies. Eur J Hum Genet 2009; 17: 749–758.

16 Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-Martin C, Walsh T et al. Strong association of de novo copy number mutations with autism. Science 2007; 316: 445–449.

17 Kusenda M, Sebat J. The role of rare structural variants in the genetics of autism spectrum disorders. Cytogenet Genome Res 2008; 123: 36–43.

18 Losh M, Sullivan PF, Trembath D, Piven J. Current developments in the genetics of autism: from phenome to genome. J Neuropathol Exp Neurol 2008; 67: 829–837.

19 Buizer-Voskamp JE, Franke L, Staal WG, van Daalen E, Kemner C, Ophoff RA et al. Systematic genotype-phenotype analysis of autism susceptibility loci implicates additional symptoms to co-occur with autism. Eur J Hum Genet 2010; 18: 588–595.

20 Gilman SR, Iossifov I, Levy D, Ronemus M, Wigler M, Vitkup D. Rare de novo variants associated with autism implicate a large functional network of genes involved in formation and function of synapses. Neuron 2011; 70: 898–907.

21 Sanders SJ, Ercan-Sencicek AG, Hus V, Luo R, Murtha MT, Moreno-De-Luca D et al. Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. Neuron 2011; 70: 863–885.

22 Geschwind DH. Autism: many genes, common pathways? Cell 2008; 135: 391–395.

23 Freitag CM. The genetics of autistic disorders and its clinical relevance: a review of the literature. Mol Psychiatry 2007; 12: 2–22.

24 Neale BM, Kou Y, Liu L, Ma'ayan A, Samocha KE, Sabo A et al. Patterns and rates of exonic de novo mutations in autism spectrum disorders. Nature 2012; 485: 242–245.

25 Geschwind DH, Sowinski J, Lord C, Iversen P, Shestack J, Jones P et al. The autism genetic resource exchange: a resource for the study of autism and related neuropsychiatric conditions. Am J Hum Genet 2001; 69: 463–466.

26 Guzzetta G, Jurman G, Furlanello C. A machine learning pipeline for quantitative phenotype prediction from genotype data. BMC Bioinformatics 2010; 11(Suppl 8): S3.

27 Lord C, Rutter M, Le Couteur A. Autism Diagnostic interview-revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. J Autism Dev Disord 1994; 24: 659–685.

28 International HapMap Consortium. The International HapMap Project. Nature 2003; 426: 789–796.

29 Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007; 81: 559–575.

30 Perez-Perez JM, Candela H, Micol JL. Understanding synergy in genetic interactions. Trends Genet 2009; 25: 368–376.

31 Mosconi MW, Kay M, D'Cruz AM, Guter S, Kapur K, Macmillan C et al. Neurobehavioral abnormalities in first-degree relatives of individuals with autism. Arch Gen Psychiatry 2010; 67: 830–840.

32 Cavalleri GL, Weale ME, Shianna KV, Singh R, Lynch JM, Grinton B et al. Multicentre search for genetic susceptibility loci in sporadic epilepsy syndrome and seizure types: a case-control study. Lancet Neurol 2007; 6: 970–980.

33 Lee US, Cui J. {beta} subunit-specific modulations of BK channel function by a mutation associated with epilepsy and dyskinesia. J Physiol 2009; 587: 1481–1498.

34 Monk CS, Weng SJ, Wiggins JL, Kurapati N, Louro HM, Carrasco M et al. Neural circuitry of emotional face processing in autism spectrum disorders. J Psychiatry Neurosci 2010; 35: 105–114.

35 Jiang M, Gold MS, Boulay G, Spicher K, Peyton M, Brabet P et al. Multiple neurological abnormalities in mice deficient in the G protein Go. Proc Natl Acad Sci USA 1998; 95: 3269–3274.

36 Masuho I, Mototani Y, Sahara Y, Asami J, Nakamura S, Kozasa T et al. Dynamic expression patterns of G protein-regulated inducer of neurite outgrowth 1 (GRIN1) and its colocalization with Galphao implicate significant roles of Galphao-GRIN1 signaling in nervous system. Dev Dyn 2008; 237: 2415–2429.

37 Yang H, Wan L, Song F, Wang M, Huang Y. Palmitoylation modification of Galpha(o) depresses its susceptibility to GAP-43 activation. Int J Biochem Cell Biol 2009; 41: 1495–1501.

38 Zikopoulos B, Barbas H. Changes in prefrontal axons may disrupt the network in autism. J Neurosci 2010; 30: 14595–14609.

39 Toal F, Daly EM, Page L, Deeley Q, Hallahan B, Bloemen O et al. Clinical and anatomical heterogeneity in autistic spectrum disorder: a structural MRI study. Psychol Med 2010; 40: 1171–1181.

40 Drouin-Ouellet J, Brownell AL, Saint-Pierre M, Fasano C, Emond V, Trudeau LE et al. Neuroinflammation is associated with changes in glial mGluR5 expression and the development of neonatal excitotoxic lesions. Glia 2011; 59: 188–199.

41 Le Duigou C, Holden T, Kullmann DM. Short- and long-term depression at glutamatergic synapses on hippocampal interneurons by group I mGluR activation. Neuropharmacology 2011; 60: 748–756.

42 Popkirov SG, Manahan-Vaughan D. Involvement of the metabotropic glutamate receptor mGluR5 in NMDA receptor-dependent, learning-facilitated long-term depression in CA1 synapses. Cereb Cortex 2011; 21: 501–509.

43 Suzuki E, Okada T. Group I metabotropic glutamate receptors are involved in TEA-induced long-term potentiation at mossy fiber-CA3 synapses in the rat hippocampus. Brain Res 2010; 1313: 45–52.

44 Fowler SW, Ramsey AK, Walker JM, Serfozo P, Olive MF, Schachtman TR et al. Functional interaction of mGlu5 and NMDA receptors in aversive learning in rats. Neurobiol Learn Mem 2011; 95: 73–79.

45 Sacco R, Curatolo P, Manzi B, Militerni R, Bravaccio C, Frolli A et al. Principal pathogenetic components and biological endophenotypes in autism spectrum disorders. Autism Res 2010; 3: 237–252.

46 Boyd BA, Baranek GT, Sideris J, Poe MD, Watson LR, Patten E et al. Sensory features and repetitive behaviors in children with autism and developmental delays. Autism Res 2010; 3: 78–87.

47 Byrnes KR, Stoica B, Loane DJ, Riccio A, Davis MI, Faden AI. Metabotropic glutamate receptor 5 activation inhibits microglial associated inflammation and neurotoxicity. Glia 2009; 57: 550–560.

48 Vargas DL, Nascimbene C, Krishnan C, Zimmerman AW, Pardo CA. Neuroglial activation and neuroinflammation in the brain of patients with autism. Ann Neurol 2005; 57: 67–81.

49 Morgan JT, Chana G, Pardo CA, Achim C, Semendeferi K, Buckwalter J et al. Microglial activation and increased microglial density observed in the dorsolateral prefrontal cortex in autism. Biol Psychiatry 2010; 68: 368–376.

50 Pantelis C, Velakoulis D, McGorry PD, Wood SJ, Suckling J, Phillips LJ et al. Neuroanatomical abnormalities before and after onset of psychosis: a cross-sectional and longitudinal MRI comparison. Lancet 2003; 361: 281–288.

Supplementary information accompanies the paper on the Molecular Psychiatry website (http://www.nature.com/mp)

Author/s:
Skafidas, E;Testa, R;Zantomio, D;Chana, G;Everall, IP;Pantelis, C

Title:
Predicting the diagnosis of autism spectrum disorder using gene pathway analysis

Date:
2014-04-01