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 (KCNM8, 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.
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
Predicting ASD phenotype based upon candidate SNPs

For each individual, a 775-dimensional vector was constructed, corresponding to the 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. 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).

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 ADCYS 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 33.5% 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                   | $5.0 \times 10^{-7}$       | $5.0 \times 10^{-7}$       |
| hsa04540      | Gap junction                        | $5.0 \times 10^{-7}$       | $5.0 \times 10^{-7}$       |
| hsa04730      | Long-term depression                | $5.0 \times 10^{-7}$       | $5.0 \times 10^{-7}$       |
| hsa04070      | Phosphotidylinositol signaling      | $1.5 \times 10^{-6}$       | $5.0 \times 10^{-7}$       |
| hsa04720      | Long-term potentiation              | $2.5 \times 10^{-6}$       | $5.0 \times 10^{-7}$       |
| hsa00230      | Purine metabolism                   | $1.0 \times 10^{-5}$       | $5.0 \times 10^{-7}$       |
| hsa04010      | Mitogen-activated kinase-like protein | $5.0 \times 10^{-7}$     | —                         |
| hsa04740      | Olfactory transduction              | $5.0 \times 10^{-7}$       | —                         |
| hsa04910      | Insulin signaling pathway           | $1.5 \times 10^{-6}$       | —                         |
| hsa04916      | Melanogenesis                       | $2.0 \times 10^{-6}$       | —                         |
| hsa04310      | Wnt signaling                       | $4.0 \times 10^{-6}$       | —                         |
| hsa04912      | GnRH signaling                      | $4.5 \times 10^{-6}$       | —                         |
| hsa04120      | Ubiquitin-mediated proteolysis      | $7.0 \times 10^{-6}$       | —                         |
| hsa04080      | Neuroactive ligand receptor         | $1.2 \times 10^{-5}$       | $5.0 \times 10^{-7}$       |
| hsa04062      | Chemokine signaling pathway         | $1.2 \times 10^{-5}$       | $5.0 \times 10^{-7}$       |
| hsa04060      | Cytokine–cytokine receptor          | $1.65 \times 10^{-5}$      | $5.0 \times 10^{-7}$       |
| hsa04114      | Oocyte meiosis                      | —                          | $5.0 \times 10^{-7}$       |
| hsa04360      | Axon guidance                       | —                          | $5.0 \times 10^{-7}$       |
| hsa04510      | Focal adhesion                      | —                          | $5.0 \times 10^{-7}$       |
| hsa04514      | Cell adhesion molecules             | —                          | $5.0 \times 10^{-7}$       |
| hsa04670      | Leukocyte transendothelial migration| —                          | $5.0 \times 10^{-7}$       |
| hsa04144      | Endocytosis                         | —                          | $2.0 \times 10^{-6}$       |
| hsa04742      | Taste transduction                  | —                          | $2.0 \times 10^{-6}$       |

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
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-CU individuals fell as distinct groups between the ASD and controls, further reinforcing a genetic basis for ASD with neurobehavioral abnormalities reported in parents of ASD individuals also supporting our findings. 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 variables 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. 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. 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. GNAO1 has also been shown to have a role in neuronal development co-localizing with GRIN1 at neuronal dendrites and synapses, and interacting with GAP-43 at neuronal growth cones, with increased levels of GAP-43 demonstrated in the white matter adjacent to the anterior cingulate cortex in brains from ASD patients.

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. GRM5 has a role in synaptic plasticity, modulation of synaptic excitation, innate immune function and microglial activation. GRM5-positive allosteric modulators can reverse the negative behavioral effects of NMDA receptor antagonists, including stereotypes, sensory
motor gating deficits and deficits in working, spatial and recognition memory. Features described in ASD.44–46 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, while 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.

ACKNOWLEDGEMENTS
Professor Christos Pantelis was supported by a NHMRC Senior Principal Research Fellowship (ID 628366). AGRE: 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). SPARR: We are grateful to all of the families at the participating Simons Simplex Collection (SSC) sites, as well as the principal investigators (A Beaumont, R Bernard, J Constantino, E Cook, E Fombonne, D Geschwind, R Goin-Kochel, E Hanson, D Grice, A Klin, D Ledbetter, C Lord, C Martin, D Martin, R Maxim, J Miles, O Ousley, K Pelphrey, B Peterson, J Piggott, C Saulnier, M State, W Stone, J Sutcliffe, C Walsh, Z Warren, E Wijsman). WTBC: We acknowledge use of the British 1958 Birth Cohort DNA collection, funded by the Medical Research Council grant G01234567 and the Welcome Trust grant 012345.

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 ed. 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.
34 Monk CS, Weng SJ, Wiggins JL, Kurapatii N, Lourou 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 Galphao 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. Glio 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: 746–756.

42 Popkrov 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, Sperfozo 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, Millietni 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. Glio 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.

This work is licensed under the Creative Commons Attribution-NonCommercial-No Derivative Works 3.0 Unported License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-nd/3.0/

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