Haplotype structure enables prioritization of common markers and candidate genes in autism spectrum disorder

BN Vardarajan1,2, A Eran3,4, J-Y Jung1, LM Kunkel5 and DP Wall1

Autism spectrum disorder (ASD) is a neurodevelopmental condition that results in behavioral, social and communication impairments. ASD has a substantial genetic component, with 88–95% trait concordance among monozygotic twins. Efforts to elucidate the causes of ASD have uncovered hundreds of susceptibility loci and candidate genes. However, owing to its polygenic nature and clinical heterogeneity, only a few of these markers represent clear targets for further analyses. In the present study, we used the linkage structure associated with published genetic markers of ASD to simultaneously improve candidate gene detection while providing a means of prioritizing markers of common genetic variation in ASD. We first mined the literature for linkage and association studies of single-nucleotide polymorphisms, copy-number variations and multi-allelic markers in Autism Genetic Resource Exchange (AGRE) families. From markers that reached genome-wide significance, we calculated male-specific genetic distances, in light of the observed strong male bias in ASD. Four of 67 autism-implicated regions, 3p26.1, 3p26.3, 3q25-27 and 5p15, were enriched with differentially expressed genes in blood and brain from individuals with ASD. Of 30 genes differentially expressed across multiple expression data sets, 21 were within 10 cM of an autism-implicated locus. Among them, CNTN4, CADPS2, SUMF1, SLC9A9, NTRK3 have been previously implicated in autism, whereas others have been implicated in neurological disorders comorbid with ASD. This work leverages the rich multimodal genomic information collected on AGRE families to present an efficient integrative strategy for prioritizing autism candidates and improving our understanding of the relationships among the vast collection of past genetic studies.

Translational Psychiatry (2013) 3, e262; doi:10.1038/tp.2013.38; published online 28 May 2013

Introduction

Autism spectrum disorder (ASD) is a neurodevelopmental condition that results in behavioral, social and communication impairments. It is currently estimated that 1 in every 88 children in the United States is affected with ASD, with boys five times more likely to be affected than girls.1 ASD has a substantial genetic component,2–4 with 88–95% monozygotic twin concordance and an estimated heritability of 60–90%.5 A recent study showed that a large proportion of the variance in liability among monozygotic twins can be explained by shared environmental factors (55% for autism and 58% for ASD) in addition to moderate genetic heritability (37% for autism and 38% for ASD).6 Studies conclude that there are multiple genetic factors that have a role in the etiology of autism. Recent findings have provided evidence in support of roles for de novo mutations,7–10 common genetic variants,11 rare variants12 and copy-number variation.13–16 Nevertheless, the genetic basis of the majority of ASD remains largely unclear.

Contributing to the complexity, ASD linkage studies have uncovered over 70 susceptibility loci across the genome and a large number of gene candidates,16,17 but most of these findings have not been successfully replicated. The only exceptions to this trend have been linkage peaks on 17q11–17q2118–21 and 7q22–26. Yet, linkage and association studies have dominated the approaches to disentangle the genetic etiology of autism for more than two decades, leaving behind a rich legacy of research findings in the biomedical literature. Reports of significant linkage peaks represent an important clue to the genetic cause of autism that should not be ignored, even in the absence of sufficient replication. Aside from the possibility of false positives, absence of replication could be due to several factors such as lack of sample size, differential recombination rates in the replication population, lower coverage in the replication samples of genetic markers in the linkage peaks or batch effects. However, the mechanistic relevance of the marker should still be determined. For example, a marker may designate collections of genes involved in biological processes or individual genes with mutations of high importance to the susceptibility to autism. Furthermore, these markers and their importance to the etiology of autism, once they have achieved the minimum significance threshold of logarithm-of-the-odds of 3.0 or an association P-value of <0.05 (corrected for multiple testing), are usually treated as equal. Therefore, despite the fact that markers provide maps, the granularity of those maps is insufficient to direct prioritized experimental follow-up, as every marker, and every gene proximal to that marker, is
equally likely to be as important. Given that markers have been identified on nearly every chromosome, the utility of linkage studies for providing specific gene leads and directing further experimental research is limited.

In the present study, we have focused on maximizing the value of previously published linkage and association findings using families from the Autism Genetic Resource Exchange ( AGRE) project for directing further genetic analysis of autism. Specifically, our aim was to provide finer resolution to published linkage and association studies through a novel analytical strategy focused on marker-to-gene male-specific genetic distance. Our study was loosely predicated on the assumption that genes in tight linkage with a susceptibility locus are more likely to be linked with the phenotype of interest, that is, autism, and was leveraged by the collective understanding that the disorder has a substantial male bias. As such, our work focused on reconstructing the male-specific structure of linkage disequilibrium (LD) surrounding significant autism markers to sets of genes in tight, medium and distant LD with those markers. We examined the biological signal inherent to each concept and measured its expression in peripheral blood and postmortem brain tissue from individuals with autism as compared with controls. This strategy improves the resolution of marker-based findings by pointing to the specific genes contributing to the linkage and/or association signals, more likely to have a role in ASD. A large percentage of these genes had not been previously linked to autism but had been implicated in numerous other neurological diseases, including those with overlapping symptoms. Given the ability of this strategy to identify important and novel signal among the rich collection of research findings from various linkage and association studies in autism, we anticipate that it will have broader applications in the study of other complex genetic disorders in which a large collection of samples had been previously typed and not immediately available for modern sequencing.

Materials and methods

Autism marker selection. We first mined the autism literature to identify genetic studies focusing on AGRE families. Owing to the focus on AGRE families, all probands included here were assessed and diagnosed using the same instruments and procedures. We identified 67 reports of significant autism linkage and association signals spanning 18 chromosomes (Table 1). Significance thresholds were a logarithm-of-the-odds score >3, which is suggestive evidence of linkage or corrected-association P-value <0.01 (depending on the number of markers tested in the study). The search was restricted to studies performed on AGRE families because the same subjects were used to calculate the genetic map around autism markers. This strategy allowed us to capture the true rates of recombination in the studied population and avoid any potential recombination bias. As the linkage and association studies were based on various experimental designs, we developed the strategy described below to enable their meta-analysis.

Each marker was first mapped to the NCBI human genome build 36.3. Then, a 20-Mb slice flanking that genomic coordinate was retrieved and the single-nucleotide polymorphisms (SNPs) within that region were used for calculating a genetic map using the same subjects’ genotypes. The nearest SNP to the autism marker was used as the reference for calculating recombination rates with other SNPs. The recombination rates were determined with respect to the reference. We assumed that the recombination rates between the marker and the nearest SNP was negligible, enabling us to designate that SNP as a proxy for the marker. Owing to the heterogeneity in the discovery methods of the various regions (linkage vs association, copy-number variations vs SNPs and so on), we treated each region as equally significant. This enabled us to use an unbiased approach in finding genes and regions that were enriched for autism cases.

Calculation of LD structure of autism markers. In order to establish the male-specific LD structure between genes and autism markers, we created genetic maps from a 20-Mb slice of the chromosome flanking each linkage locus. Specifically, we collected and assembled SNPs 10 Mb upstream and 10 Mb downstream of each autism marker using the SNP data for AGRE probands. As autism is almost five times more prevalent in males, we filtered out the females from the data set before calculating the genetic map. These filtration procedures followed the logic that an AGRE data specific and male-only genetic map would be the most likely to provide an accurate reflection of the samples contributing to the linkage and association signals reported in the pooled studies.

To create the genetic maps for each autism marker, we estimated fine-scale recombination rates using the LDhat software package. This program estimates recombination rates between adjacent SNPs by fitting a Bayesian model based on coalescent theory to analyze patterns of LD in the data. We conducted this analysis for all 67 markers, identifying the male-specific genetic distances between the marker and genes surrounding that marker, measured in cM. For further filtering, we pruned the genetic map to 15 cM around the marker. A process flow for the creation of these LD structure (LDS) sets is depicted in Figure 1.

Messenger RNA expression data processing. Gene Expression Omnibus data sets GSE6575 and GSE28521 were used to examine the expression of genes surrounding significant autism markers in individuals with ASD. The GSE6575 data set consists of 17 samples of individuals with ASD without regression, 18 individuals with ASD with regression, 9 patients with mental retardation or developmental delay, and 12 typically developing children from the general population. In this previous study, total RNA was extracted from whole blood samples using the PaxGene (Qiagen, Germantown, MD, USA) Blood RNA System and run on Affymetrix U133plus2.0 (Santa Clara, CA, USA). For the purposes of our study, we elected to use the 35 individuals with autism and 12 control samples from the general population. Preprocessing and expression analyses were done with the Bioinformatics Toolbox Version 2.6 (for Matlab R2007a, Mathworks, Natick, MA, USA). GeneChip Robust Multi-array Average was used for background adjustment, and control probe intensities were used to estimate nonspecific binding. Housekeeping genes, gene
Abbreviations: AGRE, Autism Genetic Resource Exchange; SNP, single-nucleotide polymorphism. Linkage and association studies performed in AGRE families were compiled and genome-wide significant markers identified. The logarithm-of-the-odds (LOD) scores and/or association P-values are listed for each marker. Human genome build 36.3 was used to calibrate marker position. Male-specific genetic distances were calculated using dense SNP genotypes from the same individuals.
expression data with empty gene symbols, genes with very low absolute expression values and genes with low variance were removed from the preprocessed data set.

The GSE28521 data set consisted of postmortem brain tissue samples from 19 autism cases and 17 controls from the Autism Tissue Project, using the Illumina (San Diego, CA, USA) HumanRef-8 v3.0 expression beadchip panel. Three regions of the brain previously implicated in autism were profiled in each individual: superior temporal gyrus (also known as Brodmann’s area 41/42), prefrontal cortex (BA9) and cerebellar vermis. Raw data were formatted with log2 transformation and normalized by quantile normalization. We considered probes with detection \( P \)-value \( \leq 0.05 \) for at least half of the samples for further analysis, as described here.\(^29\)

Raw \( P \)-values were generated using limma/bioconductor package in R software (http://www.bioconductor.org/packages/2.12/bioc/html/limma.html), and Benjamini and Hochberg multiple testing correction was applied to obtain adjusted \( P \)-values.

**Gene expression profiles around common autism markers.** To examine the importance of genes at varying cM distances, and to examine the level of signal relevant to autism surrounding each autism marker individually, we treated each marker region as an independent hypothesis. We then examined the differential regulation of genes within LDS sets using the messenger RNA expression profiles described above. Our hypothesis is that genes at close genetic distances from autism markers will be more differentially regulated than genes not in LD with the autism markers.

Our tests for significant differential expression deviated from standard analyses of microarray data for the primary reason that each LDS set reflected independent, prior biological knowledge. As such, we treated each LDS set as a separate collection of hypotheses, with the number of hypotheses being tested simultaneously equivalent to the number of genes in the set. To appropriately account for this multiple testing, we adjusted the nominal \( P \)-values using the \( q \)-value calculation,\(^31\) a measurement framed in terms of the false discovery rate.\(^32\) All 67 LDS sets were investigated in this way to determine the frequencies of significant, adjusted \( P \)-values \((q<0.05)\) surrounding each autism marker.

**Disease cross-referencing.** We mined eight existing gene-disease annotation resources for genes associated with neurological disorders considered to be closely related to autism.\(^33\) Diseases included tuberous sclerosis, epilepsy, seizure disorder and many others with established behavioral similarities to ASD. The databases examined included the Genetic Association Database,\(^34\) Database of Genomic Variants (http://projects.tcag.ca/variation/), dbSNP (http://www.ncbi.nlm.nih.gov/projects/SNP/), HuGE Navigator Navigator,\(^35\) Human Gene Mutation Database (http://www.hgmd.cf.ac.uk/ac/index.php), Online Mendelian Inheritance in Man (http://www.ncbi.nlm.nih.gov/omim/), GeneCards (http://www.genecards.org/) and SNPedia (http://snpedia.com/index.php/SNPedia). Results from these resources...
were integrated to create a list of genes and associated gene characteristics, which was used for comparisons with the autism LDS genes.

Results

More than 200 genetic studies were conducted on AGRE families between 2001 and 2012. These were mined to identify 67 genome-wide significant linkage and association signals for ASD (Table 1). Common markers for autism span 18 chromosomes, all with a logarithm-of-the-odds score >3 or a corrected association $P$-value $<0.01$. These studies were based on various experimental designs, mostly using multiplex families with affected sib-pairs. We calibrated the positions of significant markers using NCBI human genome build 36.3 (NCBI), and then aggregated all SNPs within a 10-MB window on either side of the marker to calculate the male-specific structure of LD around each marker. Examining the recombination rates in the same subjects allows us to build a population-specific genetic map, eliminating any genetic bias that might arise from considering ethnicity-matched controls.

Our calculations of recombination rates and LD between SNPs and common autism markers identified a total of 1426 genes within 25cM of the markers. Of those, 697 protein-coding genes were within 5cM, 450 between 5 and 10cM and 212 between 10 and 15cM from the nearest autism locus (Figure 2). Both recombination rates and gene densities varied extensively among autism markers ($28.1 \pm 7.3$ cM in the 20-Mb region around markers, spanning $35.4 \pm 10.4$ genes). There was a strong correlation ($\rho = 0.7$) between the size of the genetic map and the proportion of genes at distances $>10$cM. The highest density of genes was around RFWD2 and PAPPA2 on chromosome 1, in a copy-number variation-associated region encoding 60 genes within 24cM. Forty-eight and 90% of the genes fell within 5 and 10cM, respectively, indicating that LD was well preserved with increasing distance from the autism locus. In contrast, the region around a common copy-number variations near UNQ3037 on chromosome 3 contained 73% genes at a distance greater than $>10$cM.

Previous results indicate that the information content varies by marker and genetic distance, but do not directly demonstrate whether this information is of relevance to our understanding of the genetic etiology of autism. To test directly whether specific markers and/or regions surrounding those markers are more likely to contain promising new gene leads, we examined the regulatory patterns of each LDS set independently in two expression data sets obtained from the

Figure 2  Number of genes within 20cM of significant autism markers. Genetic distances were calculated using male-only Autism Genetic Resource Exchange (AGRE) proband single-nucleotide polymorphisms (SNPs). Genes were grouped into three distance bins indicating the extent of recombination with the autism marker. The figure displays the number of genes in tight linkage with the marker, and therefore the extent of recombination around each locus.
Gene Expression Omnibus: a blood-based messenger RNA expression data from individuals with autism and controls (GSE6575) and a transcriptomic analysis of postmortem brain RNA (GSE28521). In the blood-based expression data set, although the large majority showed no change in expression, 27 marker regions (40%) contained at least one gene with significant, multiple test-corrected differential expression (Table 2). More than 50% of the genes around markers on 3p26 (del CNTN4, del UNQ3037), 3q (D3S3045–D3S1763), 2q (rs17420138) and 5p (rs10513025) were differentially expressed in whole blood from individuals with ASD. In all, 79 genes were significantly enriched at \( q < 0.05 \) across all the marker sets out of which 31 (39%) and 60 (76%) genes lie within 5 and 10 cM of the nearest autism marker, respectively, further supporting the notion that the genes proximal to the markers represent more viable autism gene leads than genes further away.

Table 2 Differential expression of genes around common autism markers

| Marker | Number of surrounding genes with expression data | % of genes significantly differentially expressed in individuals with ASD | Number of surrounding genes with expression data | % of genes significantly differentially expressed in individuals with ASD |
|--------|-----------------------------------------------|---------------------------------------------------------------------|-----------------------------------------------|---------------------------------------------------------------------|
| del CNTN4 | 16 | 100.0 | 12 | 100.0 |
| rs10513025 | 12 | 75.0 | 12 | 100.0 |
| D3S3045–D3S1763 | 22 | 50.0 | 26 | 100.0 |
| (CNV: UNQ3037) | 19 | 89.5 | 21 | 66.7 |
| rs11656560–rs10520676 | 21 | 28.6 | 18 | 100.0 |
| del NRXN1 (Glessner et al.37) | 35 | 5.7 | 27 | 100.0 |
| del NRXN1 (Bucan et al.31) | 32 | 6.3 | 24 | 100.0 |
| rs6883048 | 43 | 4.7 | 30 | 100.0 |
| rs17088473 | 13 | 0.0 | 17 | 100.0 |
| rs4307059 | 16 | 0.0 | 7 | 100.0 |
| rs189671 = rs10038113 | 15 | 0.0 | 7 | 100.0 |
| rs7794745 | 27 | 0.0 | 23 | 100.0 |
| rs2710102 | 30 | 0.0 | 25 | 100.0 |
| rs738707 | 30 | 0.0 | 0.0 | 100.0 |
| rs344781 | 28 | 0.0 | 16 | 100.0 |
| D7S543 | 29 | 0.0 | 21 | 95.2 |
| rs1861972 = rs1861973 | 27 | 0.0 | 19 | 94.7 |
| del GRIK4 | 24 | 13.7 | 17 | 94.1 |
| del B3RAP1 | 19 | 0.0 | 14 | 92.9 |
| rs17599165 | 24 | 0.0 | 13 | 84.6 |
| rs1912960 | 24 | 0.0 | 13 | 84.6 |
| rs17599416 | 24 | 0.0 | 13 | 84.6 |
| rs757415 = rs12603112 | 25 | 4.0 | 22 | 81.8 |
| dup NLG1 | 22 | 0.0 | 16 | 81.3 |
| D1S2180 | 27 | 0.0 | 21 | 76.2 |
| Chr2:20444539–20446116 | 25 | 36.0 | 10 | 70.0 |
| LD block | | | | |
| rs1807984 | 25 | 0.0 | 0.0 | 100.0 |
| D1S1656 | 42 | 0.0 | 36 | 66.7 |
| del MDGA2 | 17 | 0.0 | 11 | 63.6 |
| FEOBACAA1-ZG03v | 30 | 0.0 | 22 | 63.6 |
| rs12740310–rs3737296–rs13410279 | 28 | 0.0 | 21 | 61.9 |
| D2S335 | 24 | 0.0 | 18 | 61.1 |
| rs4519482 | 28 | 0.0 | 18 | 61.1 |
| FEOBACAA2-ZD06v | 24 | 0.0 | 18 | 61.1 |
| rs38841 | 22 | 9.1 | 23 | 80.9 |
| rs1856830 | 22 | 9.1 | 23 | 80.9 |
| dup GUSBP5 | 17 | 0.0 | 12 | 58.8 |
| rs723477 | 9 | 11.1 | 16 | 56.3 |
| del PARK2 | 35 | 11.4 | 20 | 50.4 |
| del dup RFWD2–PAPPA2 | 32 | 0.0 | 24 | 50.0 |
| rs6826863 = rs17088473 | 12 | 0.0 | 10 | 50.0 |
| rs16599397 = rs208888 | 13 | 0.0 | 18 | 50.0 |
| dup UBE2A | 14 | 7.1 | 13 | 46.2 |
| maternal dup 15q11–13 | 15 | 6.7 | 13 | 46.2 |
| GABRB3 155CA-2 | 15 | 6.7 | 13 | 46.2 |
| rs25409 | 15 | 6.7 | 13 | 46.2 |
| dup 15q11–13 | 15 | 6.7 | 13 | 46.2 |
| rs4141463 | 10 | 20.0 | 18 | 44.4 |
| D21S1437 | 10 | 0.0 | 7 | 42.9 |
| del dup 15p11.2 | 15 | 0.0 | 12 | 41.7 |
| dup 15p13.3 BP4–BP5 | 23 | 8.7 | 19 | 31.6 |
| rs11552328 = rs6596189 | 23 | 0.0 | 16 | 31.3 |
| rs1358054 | 29 | 31.0 | 26 | 30.8 |
| rs1349313 | 24 | 0.0 | 13 | 30.8 |
| rs722628 | 24 | 0.0 | 13 | 30.8 |
| D17S1299 | 20 | 0.0 | 20 | 30.0 |
| rs536981 | 33 | 0.0 | 25 | 28.0 |
| del OR4M2–OR4M | 10 | 10.0 | 11 | 27.3 |
| del LOC605137 | 10 | 10.0 | 11 | 27.3 |
| rs2421826 | 26 | 46.2 | 23 | 26.1 |
| D17S1294–D17S1800 | 21 | 0.0 | 17 | 17.6 |
| D17S1534 | 21 | 0.0 | 17 | 17.6 |
| rs6590109 | 0.0 | 18 | 11.1 |
| rs13193457 | 27 | 0.0 | 21 | 7.5 |
| rs17420138 | 17 | 0.0 | 17 | 10.0 |
| rs1445442 | 24 | 0.0 | 0.0 | 0.0 |
| del MADDAM1 | 8 | 0.0 | 5 | 0.0 |

**Abbreviations:** ASD, autism spectrum disorder; CNV, copy-number variation. For each marker region, the table lists the percentage of genes found to be significantly differentially expressed in blood and brain of individuals with ASD at a significance level of \( q < 0.05 \).

Discussion

Despite the high heritability of autism, efforts to identify its genetic causes have enjoyed only limited success. Numerous susceptibility loci have been identified, yet few have been replicated, supporting the notion that the genetic complexity of this disorder outmatches the proportion of the population with autism that has been sampled to date. Until the sampling adequately covers the diversity of genetic systems underlying ASD, we must develop analytical approaches to make optimal use of existing results. To this end, we focused here on...
Table 3 Top candidate genes based on integrating a decade of genome-wide linkage and association studies, the autism male bias and differential expression in brain and blood of individuals with autism spectrum disorder (ASD)

| Gene          | Differential expression in blood | Differential expression in brain | Male-specific genetic distance from marker (cM) | Association with disorders comorbid to ASD |
|---------------|----------------------------------|----------------------------------|-----------------------------------------------|------------------------------------------|
| TRIM44        | 9.5e - 02                        | 5.1e - 03                        | 1.4e - 02                                     | 0.84                                     |
| ITPR1         | 7.7e - 01                        | 9.4e - 04                        | 1.5e - 03                                     | 1.09                                     |
| NHERF2        | 3.0e - 02                        | 4.5e - 02                        | 3.2e - 01                                     | 1.5e - 03                                 |
| CNTN4         | 1.4e - 01                        | 3.6e - 02                        | 4.5e - 02                                     | 1.5e - 03                                 |
| MNAT3         | 2.0e - 01                        | 4.5e - 02                        | 6.5e - 01                                     | 1.5e - 03                                 |
| RAB6B         | 2.0e - 01                        | 4.5e - 02                        | 6.5e - 01                                     | 1.5e - 03                                 |
| CADPS2        | 4.1e - 04                        | 3.7e - 03                        | 1.0e - 05                                     | 1.5e - 03                                 |
| SPTBN1        | 1.5e - 03                        | 9.1e - 03                        | 1.6e - 01                                     | 1.5e - 03                                 |
| TMEM108       | 2.0e - 01                        | 4.5e - 02                        | 1.0e - 01                                     | 1.5e - 03                                 |
| ACPL2         | 8.3e - 02                        | 3.2e - 02                        | 3.6e - 02                                     | 1.5e - 03                                 |
| ADCY2         | 1.4e - 01                        | 1.1e - 02                        | 7.7e - 01                                     | 1.5e - 03                                 |
| NSUN2         | 1.1e - 02                        | 3.2e - 02                        | 7.7e - 01                                     | 1.5e - 03                                 |
| PANK1         | 1.4e - 02                        | 3.8e - 02                        | 7.7e - 03                                     | 1.5e - 03                                 |
| SUMF1         | 1.4e - 01                        | 2.4e - 02                        | 4.9e - 01                                     | 1.5e - 03                                 |
| TANC1         | 1.2e - 01                        | 1.7e - 03                        | 6.0e - 02                                     | 1.5e - 03                                 |
| SLC23A4       | 1.6e - 02                        | 2.7e - 02                        | 2.4e - 02                                     | 1.5e - 03                                 |
| EPB41L5       | 2.5e - 03                        | 1.3e - 02                        | 3.5e - 02                                     | 1.5e - 03                                 |
| ALKBH3        | 1.6e - 01                        | 3.7e - 02                        | 3.0e - 05                                     | 1.5e - 03                                 |
| SLC9A9        | 7.5e - 02                        | 2.8e - 02                        | 4.3e - 04                                     | 1.5e - 03                                 |
| NTRK3         | 3.0e - 02                        | 4.0e - 02                        | 7.9e - 02                                     | 1.5e - 03                                 |
| PLSCR4        | 5.4e - 02                        | 2.8e - 02                        | 8.4e - 03                                     | 1.5e - 03                                 |
| MOY10         | 1.4e - 01                        | 3.2e - 02                        | 1.4e - 01                                     | 1.5e - 03                                 |
| KCCMA1        | 1.3e - 02                        | 3.8e - 02                        | 2.4e - 01                                     | 1.5e - 03                                 |
| SMYD3         | 4.8e - 04                        | 9.5e - 03                        | 2.6e - 02                                     | 1.5e - 03                                 |
| ATP2B2        | 5.3e - 02                        | 2.4e - 02                        | 1.8e - 02                                     | 1.5e - 03                                 |
| ALDH1B1A1     | 9.1e - 03                        | 3.8e - 02                        | 6.0e - 01                                     | 1.5e - 03                                 |
| LMCD1         | 1.3e - 01                        | 2.4e - 02                        | 5.4e - 01                                     | 1.5e - 03                                 |
| ATG7          | 2.7e - 01                        | 4.5e - 04                        | 3.4e - 04                                     | 1.5e - 03                                 |
| SYN2          | 2.6e - 01                        | 2.6e - 02                        | 4.6e - 03                                     | 1.5e - 03                                 |
| MKRN2         | 2.3e - 01                        | 2.6e - 02                        | 2.0e - 02                                     | 1.5e - 03                                 |

Abbreviation: FDR, false discovery rate.

Listed are genes located within 20 male-specific cM of genome-wide significant autism markers, which are also differentially expressed in both brain and blood of individuals with ASD. Of these, 19 genes (63%) were previously implicated in neurological disorders with high degrees of overlap in symptomatology and morbidity to ASD.

aList of disorders: (1) neurofibromatosis, (2) tuberous sclerosis, (3) anxiety disorders, (4) ataxia, (5) attention deficit disorder, (6) autistic disorder, (7) bipolar disorder, (8) seizures, (9) cerebral palsy, (10) dementia, (11) depressive disorder, (12) Down syndrome, (13) dystonia, (14) encephalomyelitis, (15) epilepsy, (16) schizophrenia, (17) hydrocephalus, (18) mental retardation, (19) microcephaly, (20) multiple sclerosis, (21) neuroacanthocytosis, (22) neuroaxonal dystrophies, (23) obsessive-compulsive disorder.

the development of a simple strategy aimed at targeting previously published autism markers, as well as genes genetically proximal to those markers and most likely to be causally related to ASD. By coupling the structure of LD with knowledge of biological process and patterns of gene expression data from individuals with ASD, we were able to identify a set of markers and genes proximal to those markers likely to be most informative to the genetic basis of autism. Specific loci on a few chromosomes including three signals on chromosome 3 and one on chromosome 5 yielded the greatest signal, with a sizable percentage of adjacent genes showing highly significant differential expression in blood and brain data from individuals with autism. In support of their relevance to the genetics of autism, many of the differentially expressed genes closely linked to the markers have already been identified as promising autism gene candidates, such as CNTN4, CADPS2, SUMF1, NTRK3 and SLC9A9. In addition, an even greater percentage of these genes have been linked to neurological diseases with high comorbidity and behavioral similarities to ASD.

Overall, our strategy provides a means for meta-analysis of previous linkage and association studies to prioritize both markers and adjacent genes for further experimental analysis. Although our results corroborate the general rule of thumb that genes close to loci identified via linkage and association studies are likely to be informative to the disease under study, they stress that this rule only applies to specific markers. Given the success of application to the autism research field, we expect that our analytical strategy could be of general use in the study of other similarly complex genetic diseases, such as Alzheimer’s disease and type 1 diabetes.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements. We would like to thank Autism Speaks, AGRE, and participating families for making the data for this research available. We thank Wall lab members and Professors Isaac Kohane, Marco Ramoni and Peter Tonellato for engaging discussions related to the project. This work was supported by the
Haplotype structure enables gene discovery in autism

BN Vardarajan et al

National Institute of Health Grants 1R01MH085143-01 and 1R01MH096611-01A1 awarded to DPW.

1. Autism and Developmental Disabilities Monitoring Network Surveillance Year 2008 Principal Investigators, Centers for Disease Control and Prevention. Prevalence of autism spectrum disorders—Autism and Developmental Disabilities Monitoring Network, 14 sites, United States, 2008. MMWR Surveill Summ 2012; 61: 1–19.

2. 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.

3. Lauritsen M, Pedersen E. The genetics of autism. Acta Psychiatr Scand 2001; 103: 411–427.

4. Rutter M. Genetic studies of autism: from the 1970s into the millennium. J Abnorm Child Psychol 2000; 28: 3–14.

5. Ronald A, Hoekstra R. A. Autism spectrum disorders and autistic traits: a decade of new studies. Am J Med Genet B Neuropsychiatr Genet 2011; 158B: 265–274.

6. Hallmayer J, Cleveland S, Torres A, Phillips J, Cohen B, Toriro T et al. Genetic heritability and shared environmental factors among twin pairs with autism. Arch Gen Psychiatry 2011; 68: 1095–1102.

7. Sanders SJ, Mumtah MT, Gupta AR, Murdoch JD, Rauschen MJ, Willey AJ et al. De novo mutations revealed by whole-exome sequencing are strongly associated with autism. Nature 2012; 485: 237–241.

8. Neale BM, Kuo Y, Liu M, Ma’ayan A, Samocha KE, Sabo A et al. Patterns and rates of de novo mutations in autism spectrum disorders. Nature 2012; 485: 242–245.

9. Ozsoylu I, Rombens M, Levy D, Wang Z, Hakker I, Rosenbaum J et al. De novo gene disruptions in children on the autistic spectrum. Neuron 2012; 74: 285–299.

10. Wang K, Zhang H, Duan M, Glessner J, Abrahams BS et al. Common genetic variants on 5p14.1 associate with autism spectrum disorders. Nature 2009; 459: 528–533.

11. Zhao X, Loetta 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.

12. Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-Martin C, Walsh T et al. Exome sequencing identifies a highly interconnected protein network of de novo mutations. Nature 2007; 447: 254–260.

13. Alarcon M, Cantor RM, Liu J, Gilliam TC, Geschwind DH. Evidence for a language QTL of autism and GABA receptor subunit genes in multiple ethnic groups. Mol Psychiatry 2006; 11: 215–222.

14. Buican M, Abrahams BS, Wang K, Glessner JT, Herman EI, Sonnemennik LI et al. Genome-wide analyses of exonic copy number variants in a family-based study point to novel autism susceptibility genes. PLoS Genet 2009; 5: e1000336.

15. Romoz N, Cali G, Reichert JQ, Silverman JM, Buixbaum JD. An analysis of candidate autism loci on chromosome 2q24-25: evidence for association to the HST3 K9 gene. Nat Genet 2008; 6: 1152–1158.

16. Alarcon M, Silverman JM, Smith CJ, Kilfarski M, Reichert J, Holland E et al. Evidence for a susceptibility gene for autism on chromosome 2 and for genetic heterogeneity. Am J Hum Genet 2001; 68: 1541–1550.

17. Fradin D, Cheslack-Postava K, Ladd-Acosta C, Newschaffer C, Chakravarti A, Arking D et al. Parent-of-origin effects in autism identified through genome-wide linkage analysis of 16,000 SNPs. PLoS One 2010; 5: e12513.

18. Weiss LA, Arking DE, Daly MJ, Chakravarti A. Gene Discovery Project of Jonts, H, the Autism, A. A. genome-wide linkage and association scan reveals novel loci for autism. Nature 2009; 461: 852–858.

19. Philippi A, Torres F, Carayol J, Rousseau F, Letexier M, Roschmann E et al. Association of autism with polymorphisms in the paired-like homeodomain transcription factor 1 (PITX1) on chromosome 5q13: a candidate gene analysis. BMC Med Genet 2007; 8: 74.

20. Bizzari AR, Crotobal J, Tayeb A, Merette C, Lapate A. Latent class model with familial dependence to address heterogeneity in complex diseases: adapting the approach to family-based association studies. Genet Epidemiol 2011; 35: 182–189.

21. Serafeje P, Zhong H, Mathubu Hu AH. Association of Reslin gene polymorphisms with autism. Genomics 2006; 87: 75–83.

22. Campbell DB, Sutloff JS, Ebert PJ, Mitternir R, Bravaccio C, Trillo SA et al. A variant gene that disrupts MET transcription is associated with autism. Proc Natl Acad Sci USA 2006; 103: 16834–16839.

23. Tanseme I, Nakamura M, Miyachi T, Toyota T, Yamada S, Tazai M et al. Further characterization of the autism susceptibility locus AUTS1 on chromosome 7q. Hum Mol Genet 2001; 10: 973–982.

24. Alarcon M, Cantor RM, Liu J, Gilliam TC, Geschwind DH. Evidence for a language quantitative trait locus on chromosome 7q in multiple families. Am J Hum Genet 2002; 70: 60–71.

25. Barrett S, Beck JC, Benier R, Bisson E, Braun TA, Casavant TL et al. An autosomal genomic screen for autism with evidence for linkage to a region on chromosome 7q. International Molecular Genetic Study of Autism Consortium. Hum Mol Genet 1998; 7: 571–578.

26. International Molecular Genetic Study of Autism Consortium (IMGSAC). A genome-wide screen for autism: strong evidence for linkage to chromosomes 2q, 7q, and 16p. Am J Hum Genet 2001; 69: 570–581.

27. Alarcon M, Cantor RM, Liu J, Gilliam TC, Geschwind DH. Evidence for a language quantitative trait locus on chromosome 7q in a multiplex autism families. Am J Hum Genet 2002; 70: 60–71.

28. Mcvay GA, Myers SR, Hunt S, Deloukas P, Bentley DR, Donnelly P. The fine-scale structure of recombination rate variation in the human genome. Science 2004; 304: 581–584.

29. Gregg JP, Lit L, Baron CA, Hertz-Picciotto I, Walker W, Davis RA et al. Gene expression changes in children with autism. Genomics 2008; 91: 22–29.

30. Vitekagi J, Wang X, Johnston P, Lowe JK, Tian Y, Horvath S et al. Transcriptomic analysis of autistic brain reveals convergent molecular pathology. Nature 2011; 474: 380–384.
57. Benayed R, Gharani N, Rosman I, Mancuso V, Lazar G, Kamdar S et al. Support for the homeobox transcription factor gene ENGRAILED 2 as an autism spectrum disorder susceptibility locus. Am J Hum Genet 2005; 77: 851–868.

58. Szatmari P, Paterson AD, Zwaigenbaum L, Roberts W, Brian J, Meyer KJ, Autism Genome Project. C. Mapping autism risk loci using genetic linkage and chromosomal rearrangements. Nat Genet 2007; 39: 319–328.

59. Anitha A, Nakamura K, Yamada K, Suda S, Thanseem I, Tsujii M et al. Genetic analyses of roundabout (ROBO) axon guidance receptors in autism. Am J Med Genet B Neuropsychiatr Genet 2008; 147B: 1019–1027.

60. Ma DG, Cuccaro ML, Jaworski JM, Haynes CS, Stephan DA, Parod J et al. Dissecting the locus heterogeneity of autism: significant linkage to chromosome 12q14. Mol Psychiatry 2007; 12: 376–384.

61. 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.

62. Buxbaum JD, Silverman JM, Smith CJ, Greenberg DA, Kilifarski M, Reichert J et al. Association between a GABRB3 polymorphism and autism. Mol Psychiatry 2002; 7: 311–316.

63. Delahanty RJ, Kang QJ, Brune CW, Kistner EO, Courchesne E, Cox NJ et al. Maternal transmission of a rare GABRB3 signal peptide variant is associated with autism. Mol Psychiatry 2011; 16: 86–98.

64. Miller DT, Shen Y, Weiss LA, Korn J, Anselm I, Bridgemohan C et al. Microdeletion/duplication at 15q13.2q13.3 among individuals with features of autism and other neuropsychiatric disorders. J Med Genet 2009; 46: 242–248.

65. Kumar RA, KaraMohamed S, Sudi J, Conrad DF, Brune C, Badner JA et al. Recurrent 16p11.2 microdeletions in autism. Hum Mol Genet 2008; 17: 628–638.

66. Sutcliffe JS, Delahanty RJ, Prasad HC, McCauley JL, Han Q, Jiang L et al. Allelic heterogeneity at the serotonin transporter locus (SLC6A4) confers susceptibility to autism and rigid-compulsive behaviors. Am J Hum Genet 2005; 77: 265–279.

67. Stone JL, Merriman B, Cantor RM, Yoon AL, Gilliam TC, Geschwind DH et al. Evidence for sex-specific risk alleles in autism spectrum disorder. Am J Hum Genet 2004; 75: 1117–1123.

68. Strom SP, Stone JL, Ten Bosch JR, Merriman B, Cantor RM, Geschwind DH et al. High-density SNP association study of the 17q21 chromosomal region linked to autism identifies CACNA1G as a novel candidate gene. Mol Psychiatry 2010; 15: 996–1005.

69. Campbell DB, Li C, Sutcliffe JS, Persico AM, Levitt P. Genetic evidence implicating multiple genes in the MET receptor tyrosine kinase pathway in autism spectrum disorder. Autism Res 2008; 1: 159–168.

70. Anney R, Kiel L, Pinto D, Ragan R, Conroy J, Magalhaes TR et al. A genome-wide scan for common alleles affecting risk for autism. Hum Mol Genet 2010; 19: 4072–4082.