

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

BN Vardarajan^{1,2}, A Eran^{3,4}, J-Y Jung¹, LM Kunkel⁴ 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.^{13–15} 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

17q11–17q21^{18–21} and 7q.^{22–26} 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

¹Center for Biomedical Informatics, Harvard Medical School, Boston, MA, USA; ²Gertrude H. Sergievsky Center, Columbia University, New York, NY, USA; ³Harvard-MIT Health Sciences and Technology, Cambridge, MA, USA and ⁴Division of Genetics, Program in Genomics, Boston Children's Hospital, Boston, MA, USA
Correspondence: Dr D Wall, Center for Biomedical Informatics, Harvard Medical School, 10 Shattuck Street, Boston, MA 02115, USA.
E-mail: dpwall@hms.harvard.edu

Keywords: AGRE; autism genetics; autism spectrum disorders; bibliome mining

Received 10 December 2012; revised 13 March 2013; accepted 21 April 2013

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.¹¹ 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.¹¹ 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.²⁷ 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 GSE6575²⁸ and GSE28521²⁹ 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.³⁰ Housekeeping genes, gene

Table 1 Autism markers identified in AGRE families between 2001 and 2012

Chromosome	Marker	Median marker position (bp)	Male-specific genetic map units (cM _s)	P-value/LOD (association/linkage)	References
1	dup RFWD2-PAPPA2	174522115	23.3	$P = 1.0e - 02$	37
1	rs12740310-rs3737296-rs12410279	218873645	26.7	$P = 5.0e - 04$	38
1	D1S1656	228971975	40.9	NPL = 3.21	39
1	rs6683048	235855409	37.6	$P = 2.3e - 09$	40
2	dup AK123120	13142782	41.5	$P = 3.57e - 06$	37
2	del NRXN1	50557085	20.9	$P = 3.30e - 04$	41
2	del NRXN1	51134122	21.7	$P = 4.7e - 04$	37
2	rs17420138	158585159	19.6	$P = 5.63e - 08$	40
2	rs1807984	168787136	22.9	$P = 7.0e - 03$	42
2	D2S335	172274852	22.8	HLOD = 2.99 NPL = 3.32	43
2	rs4519482	172671605	24.2	$P = 7.0e - 05$	44
2	204,444,539-204,446,116 LD block	204445327	26.6	$P = 1.8e - 06$	45
3	del CNTN4	1915556	27.0	$P = 4.7e - 04$	37
3	del UNQ3037	4218017	30.4	$P = 2.0e - 03$	37
3	D3S3045-D3S1763	138597603	23.2	Z = 3.10 $P < 1.0e - 03$	18
3	dup NLGN1	174763176	21.5	$P = 1.0e - 02$	37
4	rs17599165	46634972	18.1	$P = 1.5e - 03$	46
4	rs1912960	46648638	18.8	$P = 7.3e - 03$	46
4	rs17599416	46668195	19.1	$P = 4.0e - 03$	46
4	rs6826933-rs17088473	61133187	17.8	HLOD = 3.79 LOD = 2.96	47
4	dup GUSBP5	144850990	22.8	$P = 1e - 02$	37
5	rs10513025	9676622	38.4	$P = 1.7e - 06$	48
5	rs1896731-rs10038113	25936438	29.3	$P = 3.4e - 06$	38
5	rs4307059	26003460	31.6	$P = 3.0E4e - 08$	11
5	rs11959298-rs6596189	134395753	23.2	$P = 4.0e - 04$	49
6	rs13193457	15453984	35.0	$P = 3.0e - 05$	50
6	del PARK2	162585788	31.6	$P = 4.7e - 03$	37
7	rs736707	102917639	21.8	$P = 1.40e - 5$	51
7	rs1858830	116099675	17.0	$P = 5.0e - 06$	52
7	rs38841	116107162	16.0	$P = 6.0e - 04$	53
7	rs7794745	146120539	36.7	LOD = 3.4 $P < 2.14e - 05$	54
7	rs2710102	147205323	35.0	$P = 2.0e - 03$	55
7	D7S483	151829212	33.4	NPL = 3.7, $P = 7.9e - 05$	56
7	rs1861972-rs1861973	154946830	28.1	$P = 3.5e - 06$	57
9	rs1340513	6967633	35.7	Zlr = 3.21 $P = 7.0e - 04$	58
9	rs722628	7136888	35.7	Zlr = 3.59 $P = 6.0e - 03$	58
9	rs536861	127353265	31.6	Zlr = 3.30 $P = 5.0e - 04$	58
10	del GRID1	87945347	30.8	$P = 3.1e - 04$	37
11	rs2421826	35187181	24.5	Zlr = 3.57	58
11	rs1358054	36163248	25.6	Zlr = 3.77 $P = 8.0e - 03$	58
11	rs6590109	124264258	37.7	$P = 9.0e - 03$	59
12	rs1445442	63577561	25.0	HLOD = 4.51	60
14	del MDGA2	46796374	22.1	$P = 1.3e - 04$	41
15	del OR4M2,OR4N	19844860	26.1	$P = 9.48e - 12$	41
15	del LOC650137	19915407	24.8	$P = 9.48e - 12$	41
15	dup UBE2A	23184355	38.4	$P = 9.27e - 06$	41
15	dup 15q11-13	23704547	37.5	$P = 1.0e - 05$	37
15	maternal dup 15q11-13	23750000	36.1	p approaching 0	61
15	GABRB3 155CA-2	24559869	38.2	MTDT $P = 2.0e - 03$	62
15	rs25409	24569934	39.2	$P = 8.0e - 03$	63
15	dup 15q13 BP4-BP5	29508500	42.5	p approaching 0	64
15	rs11855650-rs10520676	77364734	23.0	HLOD = 3.09 LOD = 3.62	47
16	FE0DBACA18ZG03v	19408579	32.6	$P = 1.6e - 04$	65
16	FE0DBACA7ZD06v	24133057	26.8	$P = 1.4e - 05$	65
16	del/dup 16p11.2	30300000	45.5	$P = 1.1e - 04$	61,66
17	D17S1294-D17S1800	26183756	22.2	HLODREC = 5.8 $P = 1.59e - 07$	67
17	D17S1294	26860299	20.9	MLS = 3.2 Male-only MLS = 4.3	68
17	D17S1299	36247989	25.7	MLS = 3.6	19
17	D17S2180	44028199	24.7	MLS = 4.1	19
17	rs757415 and rs12603112	46020488	22.8	$P = 1.9e - 05$	69
17	del BZRAP1	53747037	29.0	$P = 8.0e - 04$	41
19	del MADCAM1	451915	17.8	$P = 6.0e - 04$	41
19	rs344781	48866628	23.9	$P = 6.0e - 03$	70
20	rs723477	237362	22.8	NPL LOD = 3.81	48
20	rs16999397-rs200888	958294	24.5	HLOD = 3.36 LOD = 3.38	47
20	rs4141463	14695471	34.8	$P = 3.7e - 08$	71
21	D21S1437	20568713	28.4	NPL = 3.4 $P = 3.5e - 04$	56

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.

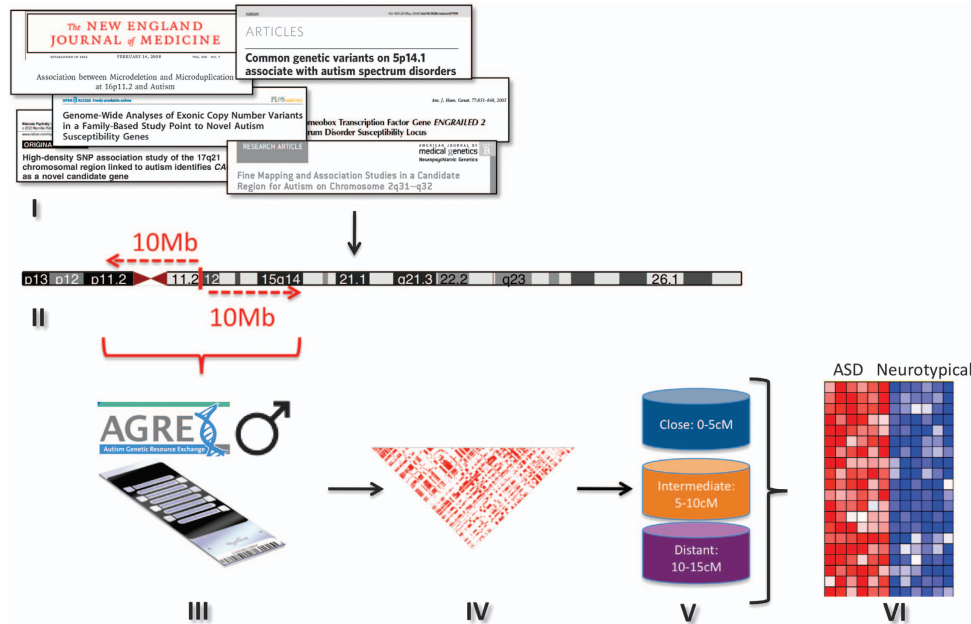


Figure 1 Integrative genomics workflow for prioritizing candidate genes for further experimentation. (I) The rich collection of genetic studies performed on Autism Genetic Resource Exchange (AGRE) families between 2001 and 2012 was mined to identify genome-wide significant linkage and association signals. (II) Markers were remapped to the current genome build (NCBI human genome build 36.3) and flanking regions extracted. (III) Single-nucleotide polymorphism (SNP) genotypes of AGRE male probands were compiled to enable male-specific genetic distance calculations in the same subjects. (IV) Regional recombination rates between markers and SNPs were calculated and (V) protein-coding genes within 20 male-specific cM from the markers identified. (VI) The expression profiles of these genes were examined in brain and blood of individuals with autism spectrum disorder (ASD) relative to neurotypical individuals. Genes found to be differentially expressed in both tissues and located within the male-specific vicinity of a significant autism marker are considered prime candidates for further studies. Of 30 genes that satisfy these criteria, 19 were previously implicated in disorders that share symptoms and morbidity patterns with ASD.

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 \log_2 transformation and normalized by quantile normalization. We considered probes with detection P -value < 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 25 cM of the markers. Of those, 697 protein-

coding genes were within 5 cM, 450 between 5 and 10 cM and 212 between 10 and 15 cM from the nearest autism locus (Figure 2). Both recombination rates and gene densities varied extensively among autism markers (28.1 ± 7.3 cM in the 20-Mb region around markers, spanning 35.4 ± 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 RFW2 and PAPA2 on chromosome 1, in a copy-number variation-associated region encoding 60 genes within 24 cM. Forty-eight and 90% of the genes fell within 5 and 10 cM, 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

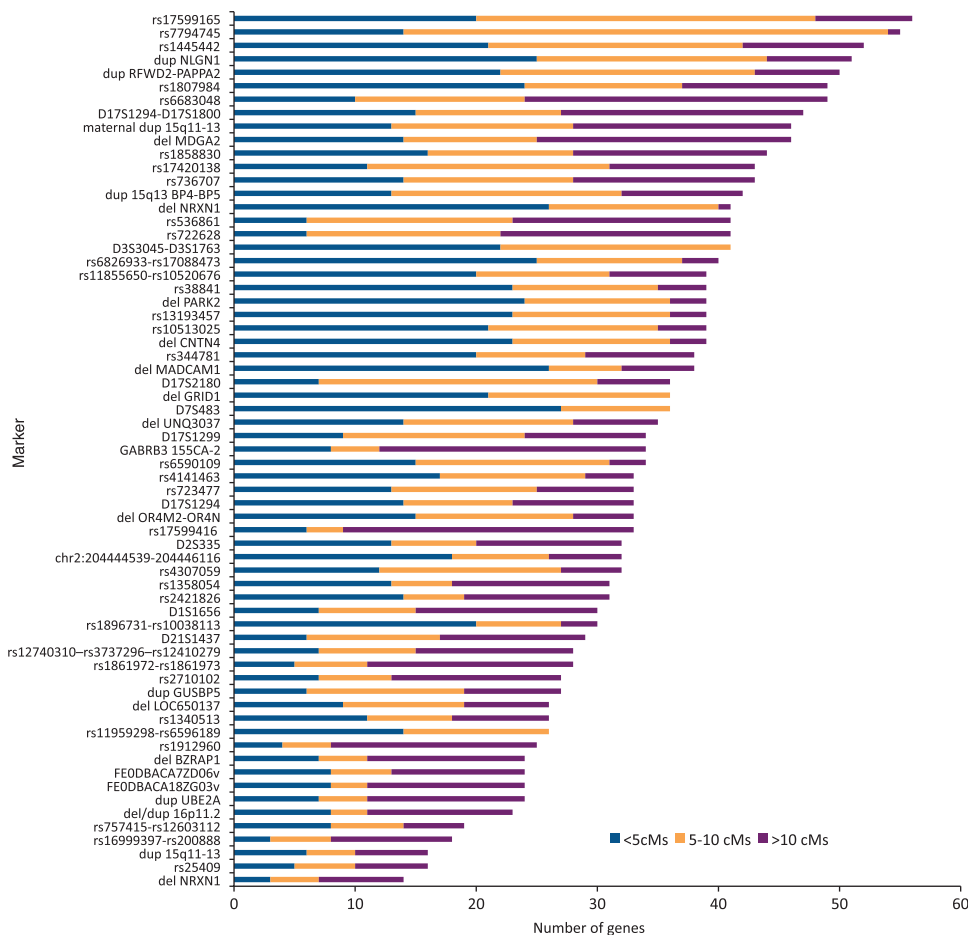


Figure 2 Number of genes within 20 cM of significant autism markers. Genetic distances were calculated using male-only Autism Genetic Resource Exchange (AGRE) proband single-nucleotide polymorphisms (SNPs).¹¹ Genes were grouped into three distance bins indicating the extent of recombination with the 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)²⁸ and a transcriptomic analysis of postmortem brain RNA (GSE28521). In the blood-based expression data set, although the large majority showed no change in

Table 2 Differential expression of genes around common autism markers

Marker	Blood (GSE6575)		Brain (GSE28521)	
	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
rs11855650–rs10520676	21	28.6	18	100.0
del NRXN1	35	5.7	27	100.0
(Glessner et al. ³⁷)				
del NRXN1 (Bucan et al. ⁴¹)	32	6.3	24	100.0
rs6683048	43	4.7	30	100.0
dup AK123120	13	0.0	17	100.0
rs4307059	16	0.0	7	100.0
rs1896731–rs10038113	15	0.0	7	100.0
rs7794745	37	0.0	23	100.0
rs2710102	38	0.0	25	100.0
rs736707	30	0.0	23	100.0
rs344781	28	0.0	16	100.0
D7S483	29	0.0	21	95.2
rs1861972–rs1861973	27	0.0	19	94.7
del GRID1	28	14.3	17	94.1
del BZRAP1	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 NLGN1	22	0.0	16	81.3
D17S2180	27	0.0	21	76.2
Chr2:204444539–204446116	25	36.0	10	70.0
LD block				
rs1807984	25	0.0	20	70.0
D1S1656	42	0.0	36	66.7
del MDGA2	17	0.0	11	63.6
FE0DBACA18ZG03v	30	0.0	22	63.6
rs12740310–rs3737296–rs12410279	28	0.0	21	61.9
D2S335	28	0.0	18	61.1
rs4519482	28	0.0	18	61.1
FE0DBACA7ZD06v	24	0.0	18	61.1
rs38841	22	9.1	23	60.9
rs1858830	22	9.1	23	60.9
dup GUSBP5	17	0.0	12	58.3
rs723477	9	11.1	16	56.3
del PARK2	35	11.4	20	50.0
dup RFW2–PAPPA2	32	0.0	24	50.0
rs6826933–rs17088473	12	0.0	10	50.0
rs16999397–rs200888	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 16p11.2	15	0.0	12	41.7
dup 15q13 BP4–BP5	23	8.7	19	31.6
rs11959298–rs6596189	23	0.0	16	31.3
rs1358054	29	31.0	26	30.8
rs1340513	24	0.0	13	30.8
rs722628	24	0.0	13	30.8
D17S1299	20	0.0	20	30.0
rs536861	33	0.0	25	28.0
del OR4M2–OR4N	10	10.0	11	27.3
del LOC650137	10	10.0	11	27.3
rs2421826	26	46.2	23	26.1
D17S1294–D17S1800	21	0.0	17	17.6
D17S1294	21	0.0	17	17.6
rs6590109	24	0.0	18	11.1
rs13193457	27	0.0	21	9.5
rs17420138	17	100.0	17	0.0
rs1445442	24	0.0	0	0.0
del MADCAM1	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 differentially expressed in blood and brain of individuals with ASD at a significance level of $q < 0.05$.

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.

In postmortem brain tissue data there was an abundance of signal in 64 of the 67 LDS sets, which contained at least one gene at q -value < 0.05 . Regions around 41 markers contained gene sets with significant differential expression, defined as $> 50\%$ of gene differentially expressed in at least one brain region between individuals with ASD and matched controls at a q -value threshold of 0.05. Of 383 genes showing evidence of differential expression at $q < 0.05$, 205 (53%) and 323 (84%) lie within 5 and 10 cM of the nearest autism marker, respectively.

Four markers were found to reside within a neighborhood of differentially expressed genes in both brain and blood of individuals with ASD. At least 50% of protein-coding genes around rs10513025, D3S3045–D3S1763, del CNTN4 and del UNQ3037 are differentially expressed in both tissues (Table 2). Three of these regions, 20 Mb around del CNTN4, del UNQ3037 and rs10513025 show heavy recombination and contain 73%, 68% and 47% of genes, respectively, at > 10 cM. Despite significant recombination within the region, genes significantly enriched for differential expression in both data sets were those closer to the autism marker. Of 30 genes found to be significantly differentially expressed in both blood and brain of individuals with ASD, 11 and 20 were within 5 and 10 cM of the nearest autism marker, respectively.

Integrating a decade of genome-wide linkage and association studies, the male bias of ASD and differential expression in both brain and blood of individuals with ASD has identified a set of 30 prime candidates for future experimentation, such as efficient targeted resequencing in very large cohorts.³⁶ Of these, CADPS2, CNTN4, NTRK3, SLC9A9 and SUMF1 have been previously implicated in ASD. Other differentially expressed genes within 20 male-specific cM of common autism markers have been implicated in disorders with shared symptoms and morbidity patterns, but have not yet been implicated in ASD (Table 3).

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 ^a
	t-test P	FDR	t-test P	FDR		
TRIM44	9.5e-02	4.8e-02	5.1e-03	1.4e-02	0.84	
ITPR1	7.7e-01	9.3e-04	1.9e-03	1.5e-03	1.09	4, 7, 15, 16
IREB2	2.3e-02	4.0e-02	1.2e-02	2.8e-03	1.39	4, 10
CNTN4	1.4e-01	2.4e-02	3.2e-01	5.1e-03	1.88	4, 6, 8, 16, 18, 19
NMNAT3	2.0e-01	4.5e-02	4.6e-01	8.3e-03	2.39	10
RAB6B	2.0e-01	4.5e-02	1.0e-03	1.9e-04	3.02	
CADPS2	4.1e-04	3.7e-03	1.0e-05	4.9e-05	3.34	5, 6
SPTBN1	1.5e-03	9.1e-03	1.6e-01	3.1e-06	3.56	1, 16
TMEM108	2.1e-01	4.5e-02	1.0e-01	2.8e-03	4.09	
ACPL2	8.3e-02	2.8e-02	8.2e-02	2.4e-03	4.43	
ADCY2	1.4e-01	3.2e-02	3.9e-02	3.9e-03	4.77	16
NSUN2	1.1e-02	1.1e-02	7.7e-01	3.8e-02	6.62	16
PANK1	1.4e-02	3.8e-02	7.7e-03	2.8e-03	7.14	
SUMF1	1.4e-01	2.4e-02	4.9e-01	6.7e-03	7.31	9, 10, 13, 21, 22
TANC1	1.2e-01	1.7e-03	6.0e-02	3.8e-02	7.31	4, 6, 17, 19, 20
SLC23A2	1.6e-02	2.7e-02	2.4e-02	1.8e-02	8.35	
EPB41L5	2.5e-03	1.3e-02	3.5e-02	1.8e-02	8.65	
ALKBH3	1.6e-01	3.7e-02	3.0e-05	2.4e-04	9.00	
SLC9A9	7.5e-02	2.8e-02	4.3e-04	1.8e-04	9.04	5, 6, 8, 15, 18, 20
NTRK3	3.0e-02	4.0e-02	7.9e-02	9.1e-03	9.67	2, 3, 5, 6, 7, 11, 15, 16, 17, 23
PLSCR4	5.4e-02	2.8e-02	8.4e-03	5.0e-04	12.30	7, 16
MYO10	1.4e-01	3.2e-02	1.4e-01	9.2e-03	12.64	10
KCNMA1	1.3e-02	3.8e-02	2.4e-01	2.2e-02	13.86	2, 8, 10, 15, 16, 18
SMYD3	4.8e-04	9.5e-03	2.6e-02	1.8e-03	14.32	
ATP2B2	5.3e-02	2.4e-02	1.1e-03	9.4e-04	14.77	14, 16, 20
ALDH18A1	9.1e-03	3.8e-02	6.0e-01	4.1e-02	15.93	8, 10, 12, 18, 19, 20
LMCD1	1.3e-01	2.4e-02	5.4e-01	6.7e-03	16.72	
ATG7	2.7e-01	4.5e-04	3.4e-04	7.1e-04	16.78	10
SYN2	2.6e-01	2.6e-02	4.6e-03	2.9e-03	18.41	7, 8, 15, 16
MKRN2	2.3e-01	2.6e-02	2.0e-02	9.5e-03	18.70	

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

National Institute of Health Grants 1R01MH085143-01 and 1R01MH090611-01A1 awarded to DPW.

- 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.
- 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.
- Lauritsen M., Ewald H. The genetics of autism. *Acta Psychiatr Scand* 2001; **103**: 411–427.
- Rutter M. Genetic studies of autism: from the 1970s into the millennium. *J Abnorm Child Psychol* 2000; **28**: 3–14.
- Ronald A, Hoekstra R. A. Autism spectrum disorders and autistic traits: a decade of new twin studies. *Am J Med Genet B Neuropsychiatr Genet* 2011; **156B**: 255–274.
- Hallmayer J, Cleveland S, Torres A, Phillips J, Cohen B, Torigoe T et al. Genetic heritability and shared environmental factors among twin pairs with autism. *Arch Gen Psychiatry* 2011; **68**: 1095–1102.
- Sanders SJ, Murtha MT, Gupta AR, Murdoch JD, Raubeson MJ, Willsey AJ et al. De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* 2012; **485**: 237–241.
- 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.
- O'Roak BJ, Vives L, Girirajan S, Karakoc E, Krumm N, Coe BP et al. Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature* 2012; **485**: 246–250.
- Iossifov I, Ronemus 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.
- Wang K, Zhang H, Ma D, Bucan M, Glessner JT, Abrahams BS et al. Common genetic variants on 5p14.1 associate with autism spectrum disorders. *Nature* 2009; **459**: 528–533.
- 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.
- 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.
- Morrow EM, Yoo SY, Flavell SW, Kim TK, Lin Y, Hill RS et al. Identifying autism loci and genes by tracing recent shared ancestry. *Science* 2008; **321**: 218–223.
- 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.
- Abrahams BS, Geschwind DH. Advances in autism genetics: on the threshold of a new neurobiology. *Nat Rev Genet* 2008; **9**: 341–355.
- Freitag CM. The genetics of autistic disorders and its clinical relevance: a review of the literature. *Mol Psychiatry* 2007; **12**: 2–22.
- Alarcon M, Yonan AL, Gilliam TC, Cantor RM, Geschwind DH. Quantitative genome scan and Ordered-Subsets Analysis of autism endophenotypes support language QTLs. *Mol Psychiatry* 2005; **10**: 747–757.
- Cantor R M, Kono N, Duvall JA, Alvarez-Retuerto A, Stone JL, Alarcon M et al. Replication of autism linkage: fine-mapping peak at 17q21. *Am J Hum Genet* 2005; **76**: 1050–1056.
- McCauley JL, Li C, Jiang L, Olson LM, Crockett G, Gainer K et al. Genome-wide and Ordered-Subset linkage analyses provide support for autism loci on 17q and 19p with evidence of phenotypic and interlocus genetic correlates. *BMC Med Genet* 2005; **6**: 1.
- Yonan AL, Alarcon M, Cheng R, Magnusson PK, Spence SJ, Palmer AA et al. A genomewide screen of 345 families for autism-susceptibility loci. *Am J Hum Genet* 2003; **73**: 886–897.
- Bailey A, Hervas A, Matthews N, Palferman S, Wallace S, Aubin A. A full genome 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.
- International Molecular Genetic Study of Autism Consortium (IMGSAC). A genomewide screen for autism: strong evidence for linkage to chromosomes 2q, 7q, and 16p. *Am J Hum Genet* 2001; **69**: 570–581.
- International Molecular Genetic Study of Autism Consortium (IMGSAC). Further characterization of the autism susceptibility locus AUTS1 on chromosome 7q. *Hum Mol Genet* 2001; **10**: 973–982.
- Alarcon M, Cantor RM, Liu J, Gilliam TC, Geschwind DH. Evidence for a language quantitative trait locus on chromosome 7q in multiplex autism families. *Am J Hum Genet* 2002; **70**: 60–71.
- Barrett S, Beck JC, Bernier R, Bisson E, Braun TA, Casavant TL et al. An autosomal genomic screen for autism. Collaborative linkage study of autism. *Am J Med Genet* 1999; **88**: 609–615.
- McVean 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.
- 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.
- Voineagu I, 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.
- Wu Z, Irizarry RA, Gentleman R, Murillo FM, Spencer F. A model based background adjustment for oligonucleotide expression arrays. *J Amer Stat Assoc* 2004; **99**: 909–917.
- Storey JD, Tibshirani R. Statistical significance for genomewide studies. *Proc Natl Acad Sci USA* 2003; **100**: 9440–9445.
- Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Roy Statist Soc Ser B* 1995; **57**: 289–300.
- Wall DP, Esteban FJ, Deluca TF, Huyck M, Monaghan T, Velez de Mendizabal N et al. Comparative analysis of neurological disorders focuses genome-wide search for autism genes. *Genomics* 2009; **93**: 120–129.
- Becker KG, Barnes KC, Bright TJ, Wang SA. The genetic association database. *Nat Genet* 2004; **36**: 431–432.
- Yu W, Gwinn M, Clyne M, Yesupriya A, Khoury MJ. A navigator for human genome epidemiology. *Nat Genet* 2008; **40**: 124–125.
- O'Roak BJ, Vives L, Fu W, Egerton JD, Stanaway IB, Phelps IG et al. Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. *Science* 2012; **338**: 1619–1622.
- Glessner JT, Wang K, Cai G, Korvatska O, Kim CE, Wood S et al. Autism genome-wide copy number variation reveals ubiquitous and neuronal genes. *Nature* 2009; **459**: 569–573.
- Ma D, Salyakina D, Jaworski JM, Konidari I, Whitehead PL, Andersen AN et al. A genome-wide association study of autism reveals a common novel risk locus at 5p14.1. *Ann Hum Genet* 2009; **73**: 263–273.
- Buxbaum JD, Silverman J, Keddache M, Smith CJ, Hollander E, Ramoz N et al. Linkage analysis for autism in a subset families with obsessive-compulsive behaviors: evidence for an autism susceptibility gene on chromosome 1 and further support for susceptibility genes on chromosome 6 and 19. *Mol Psychiatry* 2004; **9**: 144–150.
- Lu AT, Cantor RM. Allowing for sex differences increases power in a GWAS of multiplex autism families. *Mol Psychiatry* 2012; **17**: 215–222.
- Bucan M, Abrahams BS, Wang K, Glessner JT, Herman EI, Sonnenblick 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**: e1000536.
- Ramoz N, Cai G, Reichert JG, Silverman JM, Buxbaum JD. An analysis of candidate autism loci on chromosome 2q24-q33: evidence for association to the STK39 gene. *Am J Med Genet B Neuropsychiatr Genet* 2008; **147B**: 1152–1158.
- Buxbaum JD, Silverman JM, Smith CJ, Kilifarski M, Reichert J, Hollander E et al. Evidence for a susceptibility gene for autism on chromosome 2 and for genetic heterogeneity. *Am J Hum Genet* 2001; **68**: 1514–1520.
- Liu X, Novosedik N, Wang A, Hudson ML, Cohen IL, Chudley AE et al. The DLX1 and DLX2 genes and susceptibility to autism spectrum disorders. *Eur J Hum Genet* 2009; **17**: 228–235.
- Hussman JP, Chung RH, Griswold AJ, Jaworski JM, Salyakina D, Ma D et al. A noise-reduction GWAS analysis implicates altered regulation of neurite outgrowth and guidance in autism. *Mol Autism* 2011; **2**: 1.
- Collins AL, Ma D, Whitehead PL, Martin ER, Wright HH, Abramson RK et al. Investigation of autism and GABA receptor subunit genes in multiple ethnic groups. *Neurogenetics* 2006; **7**: 167–174.
- Fradin D, Cheslack-Postava K, Ladd-Acosta C, Newschaffer C, Chakravarti A, Arking D. E. et al. Parent-of-origin effects in autism identified through genome-wide linkage analysis of 16,000 SNPs. *PLoS One* 2010; **5**: e12513.
- Weiss LA, Arking DE, Daly MJ, Chakravarti A. Gene Discovery Project of Johns, H., the Autism, C. A genome-wide linkage and association scan reveals novel loci for autism. *Nature* 2009; **461**: 802–808.
- Philippi A, Tores 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 5q31: a candidate gene analysis. *BMC Med Genet* 2007; **8**: 74.
- Bureau A, Croteau J, Tayeb A, Merette C, Labbe 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.
- Serajee FJ, Zhong H, Mahbulul Huq AH. Association of Reelin gene polymorphisms with autism. *Genomics* 2006; **87**: 75–83.
- Campbell DB, Sutcliffe JS, Ebert PJ, Militeri R, Bravaccio C, Trillo S et al. A genetic variant that disrupts MET transcription is associated with autism. *Proc Natl Acad Sci USA* 2006; **103**: 16834–16839.
- Thanseem I, Nakamura K, Miyachi T, Toyota T, Yamada S, Tsujii M et al. Further evidence for the role of MET in autism susceptibility. *Neurosci Res* 2010; **68**: 137–141.
- Arking DE, Cutler DJ, Brune CW, Teslovich TM, West K, Ikeda M et al. A common genetic variant in the neurexin superfamily member CNTNAP2 increases familial risk of autism. *Am J Hum Genet* 2008; **82**: 160–164.
- 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.
- Molloy CA, Keddache M, Martin LJ. Evidence for linkage on 21q and 7q in a subset of autism characterized by developmental regression. *Mol Psychiatry* 2005; **10**: 741–746.

57. Benayed R, Gharani N, Rossman 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 DQ, Cuocaro 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 JQ, 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**: 86–96.
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. Philippi A, Roschmann E, Tores F, Lindenbaum P, Benajou A, Germain-Leclerc L *et al*. Haplotypes in the gene encoding protein kinase c-beta (PRKCB1) on chromosome 16 are associated with autism. *Mol Psychiatry* 2005; **10**: 950–960.
66. 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.
67. 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.
68. Stone JL, Merriman B, Cantor RM, Yonan 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.
69. 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.
70. 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.
71. Anney R, Klei L, Pinto D, Regan 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.



Translational Psychiatry is an open-access journal published by **Nature Publishing Group**. This work is licensed under a **Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License**. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/3.0/>