

# Exploiting aberrant mRNA expression in autism for gene discovery and diagnosis

Jinting Guan<sup>1,2</sup> · Ence Yang<sup>2,3</sup> · Jizhou Yang<sup>2</sup> · Yong Zeng<sup>1,2</sup> · Guoli Ji<sup>1,4</sup> · James J. Cai<sup>2,5</sup>

Received: 19 January 2016 / Accepted: 17 April 2016  
© Springer-Verlag Berlin Heidelberg 2016

**Abstract** Autism spectrum disorder (ASD) is characterized by substantial phenotypic and genetic heterogeneity, which greatly complicates the identification of genetic factors that contribute to the disease. Study designs have mainly focused on group differences between cases and controls. The problem is that, by their nature, group difference-based methods (e.g., differential expression analysis) blur or collapse the heterogeneity within groups. By ignoring genes with variable within-group expression, an important axis of genetic heterogeneity contributing to expression variability among affected individuals has been overlooked. To this end, we develop a new gene expression analysis method—aberrant gene expression analysis, based on the multivariate distance commonly used for outlier detection. Our method detects the discrepancies in gene expression

dispersion between groups and identifies genes with significantly different expression variability. Using this new method, we re-visited RNA sequencing data generated from post-mortem brain tissues of 47 ASD and 57 control samples. We identified 54 functional gene sets whose expression dispersion in ASD samples is more pronounced than that in controls, as well as 76 co-expression modules present in controls but absent in ASD samples due to ASD-specific aberrant gene expression. We also exploited aberrantly expressed genes as biomarkers for ASD diagnosis. With a whole blood expression data set, we identified three aberrantly expressed gene sets whose expression levels serve as discriminating variables achieving >70 % classification accuracy. In summary, our method represents a novel discovery and diagnostic strategy for ASD. Our findings may help open an expression variability-centered research avenue for other genetically heterogeneous disorders.

**Electronic supplementary material** The online version of this article (doi:[10.1007/s00439-016-1673-7](https://doi.org/10.1007/s00439-016-1673-7)) contains supplementary material, which is available to authorized users.

✉ Guoli Ji  
glji@xmu.edu.cn

✉ James J. Cai  
jcai@tamu.edu

<sup>1</sup> Department of Automation, Xiamen University, Xiamen, Fujian, China

<sup>2</sup> Department of Veterinary Integrative Biosciences, Texas A&M University, College Station, TX, USA

<sup>3</sup> Institute for Systems Biomedicine, School of Basic Medical Sciences, Peking University Health Science Center, Beijing 100191, China

<sup>4</sup> Innovation Center for Cell Signaling Network, Xiamen University, Xiamen, Fujian, China

<sup>5</sup> Interdisciplinary Program of Genetics, Texas A&M University, College Station, TX, USA

## Introduction

Autism spectrum disorder [ASD, (OMIM 209850)] is a complex neurodevelopmental condition characterized by substantial phenotypic and genetic heterogeneity (Devlin and Scherer 2012; Geschwind 2011; Geschwind and State 2015; Willsey and State 2015). Both genetic and environmental factors contribute to the increased risk of developing ASD (Persico and Bourgeron 2006; Sandin et al. 2014). Recent years have seen major advances in the understanding of the genetic, neurobiological and developmental underpinnings of ASD (Abrahams and Geschwind 2008; Belmonte et al. 2004; Elsabbagh and Johnson 2010). Genetic studies, especially genome-wide association studies (GWAS), have identified many single-nucleotide variants (SNVs) and copy number variants (CNVs) associated with ASD susceptibility

(Glessner et al. 2009; Wang et al. 2009; Weiss et al. 2009). However, it remains difficult to identify the actual causal genes underlying these associations. SNVs that produce association signals in identified loci often fall into intergenic regions, while CNVs often extend across multiple variants or genes, both of which confound the identification of causal genes. Also, there are opposing views on the relative contribution of rare versus common variants to ASD susceptibility. Some studies suggest that low-frequency variants bring a greater impact on the risk for ASD (Neale et al. 2012; Pinto et al. 2014; Sanders et al. 2012; Sebat et al. 2007), while other studies suggest that common variants form a dominating source of the risk (Gaugler et al. 2014; Klei et al. 2012). Against this background of complexity, several studies demonstrate the use of gene expression information—measuring mRNA abundance of individual genes, coupled with other genetic approaches, allows for novel insights in understanding ASD (Flint et al. 2014; Gupta et al. 2014; Voineagu et al. 2011). Analyzing gene expression and sequence data facilitates revealing the impact of regulatory genetic variants on the gene itself and the indirect consequences on the expression of other genes (Iossifov et al. 2014).

To this end, we introduce a novel, gene expression analysis method for identifying ASD-implicated genes. Our working hypothesis is that ASD is associated with aberrant gene expression caused by the promiscuous multigene activation and repression. Indeed, we show that many gene sets that contain genes known to be implicated in ASD tend to be expressed more aberrantly in ASD-affected individuals. Encouraged by these findings, we conduct a searching for unique combinations of genes for ASD diagnosis based on whole blood expression data. We use a greedy algorithm to solve the combinatorial problem of global search and identify three gene sets, each containing five genes, which can be used as classifier gene sets with high sensitivity and high specificity to discern gene expression specific to ASD patients. Altogether, our results refine the relationships between gene function and gene expression dispersion among individuals affected with ASD, providing new insights into the genetic, molecular mechanisms underlying the dysregulated gene expression in ASD. Our results also lay out the foundation for the utilization of gene expression dispersion as biomarkers for early diagnosis of ASD.

## Materials and methods

### Gene expression data

Whole transcriptomes of 104 brain tissue samples (47 ASD and 57 controls) were previously determined using

RNA sequencing by Gupta et al. (2014). The data had been deposited in the National Database for Autism Research (NDAR) under the accession code NDARCOL0002034. Among these samples, 62, 14 and 28 were tissues from cerebral cortex [Brodmann Area (BA) 19], anterior prefrontal cortex (BA 10), and a part of the frontal cortex (BA 44), respectively, resulting in 47 (32 unique individuals) ASD samples and 57 (40 unique individuals) controls. For this study, the raw data of gene expression were normalized using the conditional quantile normalization (Hansen et al. 2012) and then processed using the algorithm of probabilistic estimation of expression residuals (PEER) (Stegle et al. 2010) to remove technical variation. PEER residuals were obtained after regressing out covariates (age, gender, brain region, and sample collection site) and factors accounting for ten possible hidden determinants of expression variation. The expression median across all samples was added back to the PEER residuals to give the final processed gene expression levels. Extremely lowly expressed genes with expression median  $<2.5$  (empirical cutoff) were excluded. The final data matrix contained the expression level information of 10,127 genes in 104 samples. Principal component analysis was performed to indicate that there was no population stratification regarding the global gene expression profiles (Supplementary Figure 1).

### Functional gene sets

The curated gene sets ( $n = 10,348$ ) used in the Gene Set Enrichment Analysis (GSEA) were obtained from the molecular signatures database (MSigDB v5.0, accessed March 2015) (Liberzon et al. 2011). GO terms ( $n = 14,825$ ) associated with protein-coding genes were downloaded from BioMart (v0.7, accessed February 2015) (Smedley et al. 2015). The co-expression networks were built for control samples using the Weighted Gene Co-expression Network Analysis (WGCNA) (Langfelder and Horvath 2008). The power of 16 was chosen using the scale-free topology criterion; the minimum module size was set to 4, and the minimum height for merging modules was 0.25. The resulting modules were plotted using SBE-Toolbox (Konganti et al. 2013). The list of ASD-implicated genes was obtained from the Simons Foundation Autism Research Initiative (SFARI) Gene Scoring Module. The list includes 410 genes in the categories S and 1–4, which stand for syndromic, high confidence, strong candidate, suggestive evidence, and minimal evidence, respectively, indicating the strength of the evidence linking genes to ASD. Of the 410 genes, 294 genes are in the expression data matrix we analyzed.

## Calculation of robust MD between ASD and control samples

For a given gene set,  $MD_i$  is the Mahalanobis distance (Mahalanobis 1936) from an ASD individual  $i$  to the multivariate centroid of control individuals. Conventional  $MD_i$  was calculated using the following operation:

$$MD(x_i, x_c) = \sqrt{(x_i - x_c)^T \Psi^{-1} (x_i - x_c)}$$

where  $x_i$  is the vector of expression of genes in ASD sample  $i$ ,  $x_c$  is the vector of expression means of genes across all control samples, and  $\psi$  is the covariance matrix estimated from the controls. Throughout the study, a robust version of  $MD_i$  was calculated using the algorithm Minimum Covariance Determinant (MCD) (Rousseeuw and Van Driessen 1999). The MCD algorithm subsamples  $h$  observations out of control individuals whose covariance matrix had the smallest covariance determinant. By default,  $h = 0.75n$ , where  $n$  is the total number of control samples. The robust  $MD_i$  was then computed with the above equation by replacing  $x_c$  with the MCD estimate of location,  $\hat{x}_c$ , i.e., the expression mean of the  $h$  controls, and replacing  $\psi$  with the MCD estimate of scattering,  $(\hat{\Psi})$ , i.e., the covariance matrix estimated from the  $h$  controls. A Matlab implementation of MCD, available in the function `mc cov` of LIBRA toolbox (Verboven and Hubert 2005), was used to perform the computation of MCD estimator. For a given gene set, MCD estimates of  $(\hat{x}_c)$  and  $(\hat{\Psi})$  were computed as the outputs of `mc cov`, and re-used for calculating robust  $MD_i$  for ASD individuals.

The sum of squared  $MD_i$  ( $SSMD = \sum_{i=1}^M MD_i^2$ ) was calculated for given gene sets to measure the overall dispersion of  $M$  individuals affected with ASD. To assess the significance of SSMD of a given gene set, permutation tests were performed with  $N$  reconstructed gene sets of the same size but randomly selected genes. The  $P$  value of permutation test,  $P_{perm}$ , was determined by the ratio of  $(\frac{n}{N})$ , where  $n$  is the number of random gene sets having SSMD greater than that of the tested gene set and  $N$  is the total number of random gene sets used in permutation tests. To save computational time, we first set  $N = 1000$  to obtain a short list of gene sets with  $P_{perm} < 0.001$ , and then set  $N = 10,000$  to obtain nominal  $P_{perm}$  for gene sets in the short list. The correction for multiple testing was performed by controlling the false discovery rate (FDR) with the Benjamini–Hochberg method (Benjamini and Hochberg 1995). To measure the relative contribution of each gene in a gene set to the total SSMD of the gene set,  $\Delta SSMD$  was calculated.  $\Delta SSMD$  is the difference between the two SSMD values calculated before and after the gene is excluded from the gene set.

## Receiver operator characteristic (ROC) curve analysis

For the analysis of aberrant gene expression as the biomarker for ASD diagnosis, peripheral blood gene expression data measured using the Affymetrix Human Gene 1.0 ST array for 104 ASD and 82 controls were downloaded (GEO accession: GSE18123) (Kong et al. 2012). The raw intensity data were processed using the R function `rma` (robust multi-array average expression measure) in `Affy` package. The expression measure was quantile normalized and log2 transformed. The final data matrix contained the expression level information for 16,365 autosomal genes among 186 samples. We equally split samples into a training set and a test set, each of which contains half of ASD (i.e., 52 ASD samples) and half of control samples (i.e., 41 controls).

ROC curve analysis was used to evaluate the specificity and sensitivity of classification tests, in which gene sets were used as classifiers for ASD and controls. For a given gene set, we first obtained the multivariate centroid of controls in the training set ( $G_{training}$ ) and calculated  $MD_i$  of each sample  $i$  (including all ASD and control samples in the training set). Using ROC curve analysis, the threshold (denoted by  $T$ ) corresponding to optimal specificity and sensitivity combination was determined. If  $MD_i$  is greater than  $T$ , the sample  $i$  was classified as an ASD, otherwise a control. The performance of each gene set for predicting ASD and control samples was tested at different threshold  $T$  values to obtain the area under the curve (AUC). We denote the AUC values with respect to the training and test data sets as AUC1 and AUC2. After selecting three classifier gene sets with top AUC1 values, we then assessed their prediction performances with the test data set using ROC curve analysis again. For each classifier gene set,  $MD_i$  for all samples of the test set, regardless of the disease status of samples, was calculated against  $G_{training}$ , i.e., the multivariate centroid of controls in the training set.

## Global search for the classifier gene sets

A greedy algorithm was developed to identify subsets of genes, for which AUC1 reaches its maximal values as possible. The calculation of AUC1 can be seen in section “Receiver operator characteristic (ROC) curve analysis.” The search is global because combinations of all expressed genes were considered and no information of any predefined gene sets was used. Starting with all possible two-gene combinations, AUC1 values were computed, and the top 1000 two-gene pairs with maximal AUC1 were retained as seeds for subsequent steps. The idea of the greedy strategy is to make a locally optimal choice at each stage with the hope of finding a global optimum. Thus, the assumption here is that the genes in the two-gene combinations

producing the greatest AUC1 (i.e., the locally optimal solution) would be among those in five-gene combinations producing the greatest AUC1 (i.e., the global optimum). For each of the selected two-gene combinations, a new gene that can produce largest SSMD was identified and added to the gene pair to make a three-gene combination. The procedure was repeated until the number of genes reached five. At this stage, an additional procedure was introduced to improve the locally optimal solutions achieved by the greedy heuristic: all distinct three gene combinations were extracted from five-gene combinations as the new candidate subsets. From these three gene combinations, new genes were added to get a new round of solutions of five-gene combinations. The newly generated five-gene combinations will be retained if they produced larger AUC1 than older ones. This replacement procedure was repeated until no improvement could be made. For the top gene sets that produced the best AUC1 values, we then assessed their performances of prediction with the test data set. Computer code is available from the authors upon request.

## Results

Many biological processes underlying human diseases are accompanied by changes in gene expression in corresponding tissues (Cookson et al. 2009). ASD is not an exception. Previous studies have detected specific gene expression changes in ASD, concerning genes involved in the synaptic formation, transcriptional regulation, chromatin remodeling, or inflammation and immune response (Voineagu et al. 2011). These analyses mostly focused on departures across the average expression between the case and control groups, without considering or much less focusing on alternative patterns of departure such as those characterized by heterogeneous dispersion. The goal of the present study is to detect the difference in heterogeneous multigene expression dispersion between samples derived from ASD-affected individuals and healthy controls.

### Overview of aberrant gene expression analysis

We have previously developed a multivariate method, namely aberrant gene expression analysis, to measure the level of multigene expression dispersion in the general population (Zeng et al. 2015). This analysis method uses Mahalanobis distance (MD) to quantify the dissimilarity in multigene expression patterns between individuals (Mahalanobis 1936). MD is an appropriate measure because it accounts for the covariance between expression levels of multiple genes. The aberrant gene expression analysis can be used to identify genes more likely to be aberrantly expressed among given individuals. It can

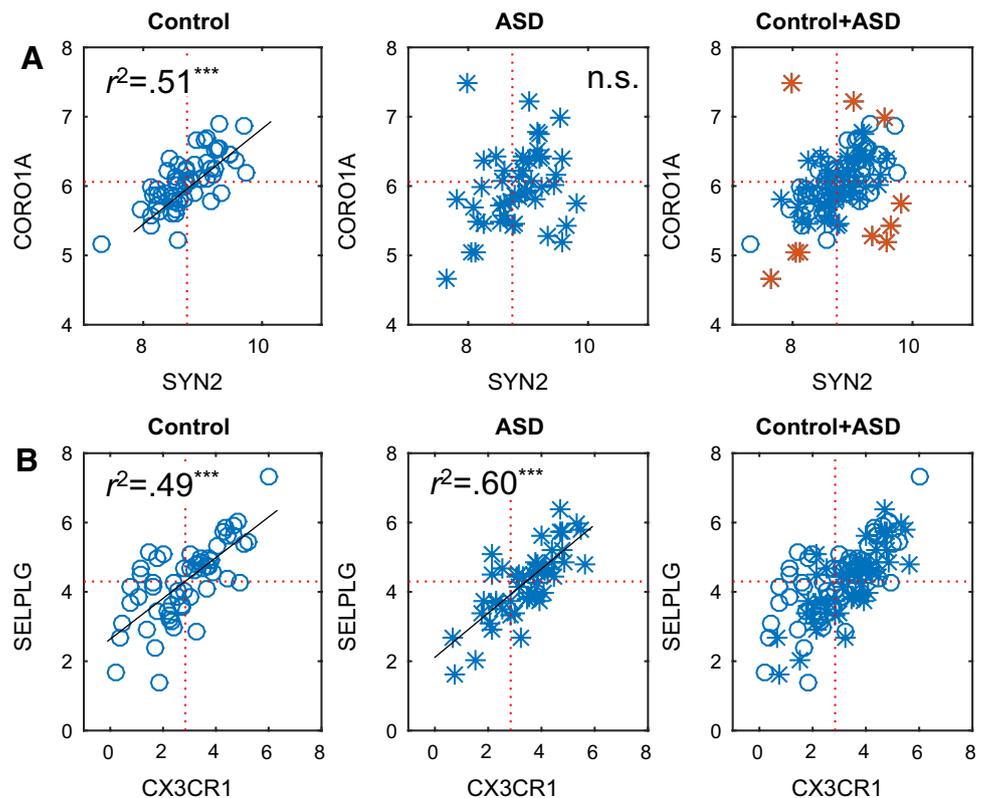
also be used to identify individual outliers whose expression for a given gene set differs markedly from most of a population.

Here, we extend the MD-based aberrant gene expression analysis to a two-group setting. We estimated the level of gene expression dispersion among individuals affected with autism relative to controls, assuming that the increased dispersion is due to a promiscuous gene activation and repression associated with autism. We applied the aberrant gene expression with such a two-group setting and re-analyzed the gene expression data generated from the post-mortem brain tissues of 47 ASD and 57 control samples (Gupta et al. 2014). For a given gene set, we computed the MD between gene expression of each ASD individual  $i$  to the multivariate centroid of the controls, denoted as  $MD_i$ . We used the sum of squared  $MD_i$  (SSMD) to measure the overall dispersion level for all ASD samples vs. the controls. Using permutation tests, we assessed the significance of gene sets and identified gene sets more likely to be aberrantly expressed among individuals affected with ASD (“Materials and methods”).

### Coordinated gene expression is disrupted in ASD

Our MD-based aberrant gene expression analysis is capable of detecting the signal of expression aberration in different forms, including, e.g., (1) the increased individual-to-individual gene expression variance (i.e., the increased gene expression variability) and (2) the decreased expression correlation between genes. To illustrate the effect of disrupted expression, we use gene sets comprising only two genes to show the aberrant gene expression among individuals affected with ASD manifested as the loss of expression correlation between the two genes. Figure 1 shows scatter plots of expression levels between gene pairs. In Fig. 1a, the expression of *CORO1A* is positively correlated with that of *SYN2* for the controls (left panel, Pearson correlation test,  $P = 3.6 \times 10^{-10}$ ). The gene *CORO1A* encodes coronin 1A, an actin binding protein. The gene *SYN2*, which is selectively expressed at nerve terminals in mature neurons, encodes synapsin II, a neuron-specific synaptic vesicle phosphoprotein (Cesca et al. 2010; Corradi et al. 2014). Synapsins interact with actin filaments in a phosphorylation-dependent manner (Benfenati et al. 1989). As evident from the description of gene functions, the correlated expression between the two genes is crucial for their respective actin-related molecular functions in normal individuals. However, such a crucial correlation becomes less significant in the ASD group (middle panel,  $P = 0.07$ ). To make the contrast, we superimposed the data points for ASD individuals onto those of the controls (right panel). The top 10 ASD samples with the largest MD values are highlighted with red asterisks. The data points of these

**Fig. 1** A proof-of-concept example, based on real data (Gupta et al. 2014), showing that **a** the correlated expression between *SYN2* and *CORO1A* presents among non-ASD samples but is disrupted among ASD samples, while **b** the correlated expression between *CX3CR1* and *SELPLG* presents among both non-ASD and ASD samples. Red stars in **a** show the top 10 ASD samples with the largest  $MD_i$  and  $r^2$  is the squared Pearson correlation coefficient



ASD samples are the most remote observations, distributed either far away from or orthogonally against the “cloud of data points” around the population mean, in which most control individuals are located. Figure 1b presents a negative example, in which the correlations in expression levels between two genes, *CX3CR1* and *SELPLG*, are presented in both control and ASD groups ( $P = 1.3 \times 10^{-9}$  and  $1.4 \times 10^{-10}$ , respectively), indicating that the coordinated expression between the two genes is not disrupted in ASD. Altogether, these two examples, one positive and one negative, suggest that aberrant gene expression is not universal. The pattern of aberration may be highly specific with regard to certain gene sets (e.g., that in Fig. 1a) but not others (e.g., that in Fig. 1b).

### Functional gene sets that tend to be aberrantly expressed in ASD

To identify gene sets more likely to be aberrantly expressed in ASD samples, we calculated SSMD for a number of pre-defined gene sets (“Materials and methods”). These included the curated gene sets in the MSigDB of GSEA (Liberzon et al. 2011). The significance of each gene set was assessed using permutation tests with random gene sets. A total of 18 GSEA curated gene sets were found to produce significantly higher SSMD than random gene sets at FDR of 10 %. Functions of these gene sets fall into

four major categories, namely, metabolism and biosynthesis, immune or inflammatory response, signaling pathway, and vitamins and supplements (Table 1). The relevance of these major functional categories with ASD is supported by respective studies (Abrahams and Geschwind 2008; Chow et al. 2012; Frye et al. 2010; Klaiman et al. 2013; Lazaro and Golshani 2015; Sawicka and Zukin 2012; Tierney et al. 2006). For example, the mTOR signaling pathway, which has a full name in the Reactome database: Energy dependent regulation of the serine/threonine protein kinase mTOR by LKB1-AMPK (Croft et al. 2014), is essential to synaptogenesis; gene products of the pathway regulate dendritic spine morphology in synapses. The dysregulation of this pathway is implicated in ASD (Abrahams and Geschwind 2008; Lazaro and Golshani 2015; Sawicka and Zukin 2012). Table 1 also contains four gene sets with miscellaneous functions, unclassified into any of the four major categories but all implicated in ASD. These genes are involved in: (1) activated point mutants of *FGFR2* (Schubert et al. 2015; Stevens et al. 2010), (2) activation of the AP-1 family of transcription factors (Schaaf et al. 2011), (3) inwardly rectifying  $K^+$  channels (Guglielmi et al. 2015; Lee et al. 2014), and (4) G2/M checkpoints (Fatemi et al. 2008).

To determine individual gene’s contribution to the total SSMD of a gene set, we computed  $\Delta$ SSMD for each gene. The  $\Delta$ SSMD of a gene is the difference between SSMD values of a gene set before and after excluding

**Table 1** GSEA curated gene sets that tend to be aberrantly expressed in ASD

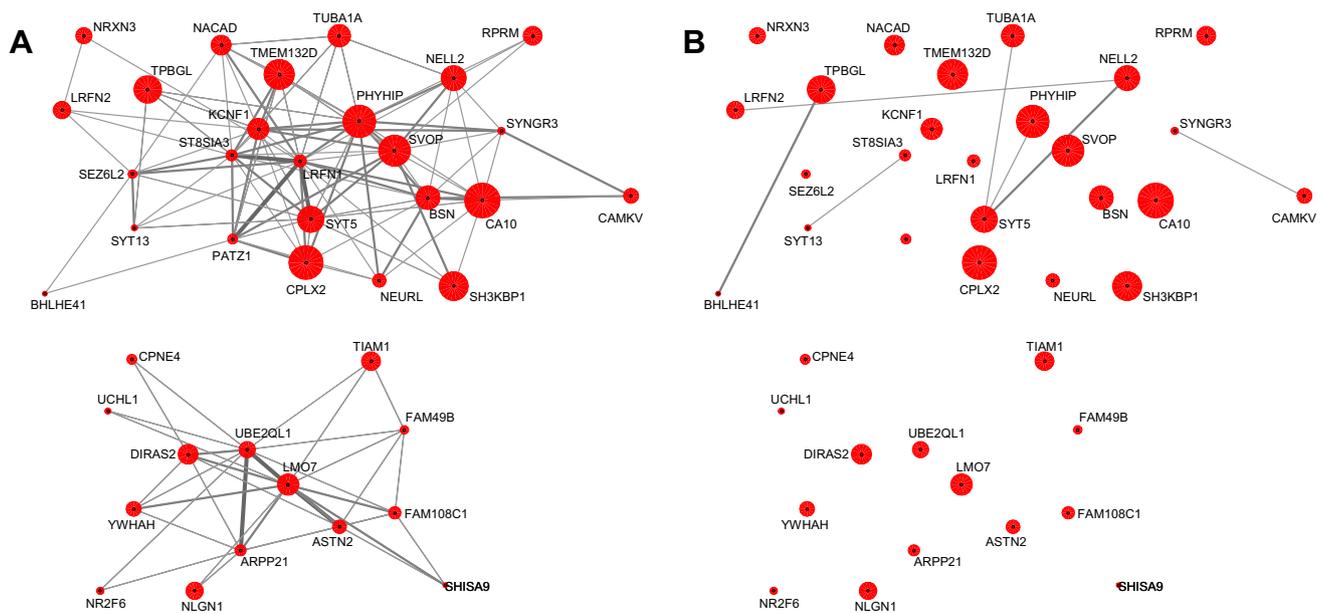
GSEA gene set	Number of genes <sup>a</sup>	Top $\Delta$ SSMD gene	References
<b>Metabolism and biosynthesis</b>			
KEGG_PENTOSE_PHOSPHATE_PATHWAY	19/27	<i>H6PD, PRPS2, PFKP</i>	
KEGG_STEROID_BIOSYNTHESIS	14/17	<i>SC5DL, NSDHL, <b>DHCR7</b></i>	
REACTOME_CHOLESTEROL_BIOSYNTHESIS	20/24	<i>SQLE, HSD17B7, HMGCR</i>	(Tierney et al. 2006)
REACTOME_BRANCHED_CHAIN_AMINO_ACID_CATABOLISM	16/17	<i>DLD, HIBADH, MCCC2</i>	
<b>Immune/Inflammatory response</b>			
BIOCARTA_LAIR_PATHWAY	4/17	<i>SELPLG, C3, ITGB1</i>	
BIOCARTA_41BB_PATHWAY	12/17	<i>MAPK8, ATF2, MAPK14</i>	
REACTOME_IL1_SIGNALING	25/39	<i>CHUK, RBX1, BTRC</i>	(Chow et al. 2012)
REACTOME_REGULATION_OF_IFNA_SIGNALING	6/24	<i>STAT1, PTPN1, JAK1</i>	
<b>Signaling pathway</b>			
BIOCARTA_IGF1_PATHWAY	20/21	<i>JUN, CSNK2A1, ELK1</i>	
PID_S1P_S1P2_PATHWAY	21/24	<i>MAPK8, MAPK14, JUN</i>	
PID_HNF3APATHWAY (FOXA1/HNF3A TF network)	22/44	<i>NDUFV3, PISD, FOS</i>	
REACTOME_ENERGY_DEPENDENT_REGULATION_OF_MTOR_BY_LKB1_AMPK	15/18	<i>PRKAA1, CAB39, <b>TSC1</b></i>	(Abrahams and Geschwind 2008; Lazaro and Golshani 2015; Sawicka and Zukin 2012)
<b>Vitamins and supplements</b>			
BIOCARTA_VITCB_PATHWAY	6/11	<i>SLC2A3, COL4A2, SLC2A1</i>	
REACTOME_TETRAHYDROBIOP-TERIN_BH4_SYNTHESIS_RECYCLING_SALVAGE_AND_REGULATION	9/13	<i>GCHFR, PTS, AKT1</i>	(Frye et al. 2010; Klaiman et al. 2013)
<b>Miscellaneous</b>			
REACTOME_ACTIVATED_POINT_MUTANTS_OF_FGFR2	4/16	<i>FGF9, FGFR2, FGF1</i>	(Schubert et al. 2015; Stevens et al. 2010)
REACTOME_ACTIVATION_OF_THE_API_FAMILY_OF_TRANSCRIPTION_FACTORS	10/10	<i>MAPK14, <b>MAPK3</b>, ATF2</i>	(Schaaf et al. 2011)
REACTOME_INWARDLY_RECTIFYING_K_CHANNELS	20/31	<i><b>KCNJ10</b>, KCNJ4, GNG4</i>	(Guglielmi et al. 2015; Lee et al. 2014)
REACTOME_G2_M_CHECKPOINTS	22/45	<i>MCM2, RFC5, RPA2</i>	(Fatemi et al. 2008)

Gene sets mentioned in the main text are shown in italic. SFARI ASD-implicated genes are shown in bold

<sup>a</sup> Number of genes included in our analysis/Number of genes in the gene set

the gene from the gene set. Top three genes with the largest  $\Delta$ SSMD are given for all gene sets in Table 1. SFARI ASD-implicated genes are highlighted. To further investigate the relationship between gene function and aberrant gene expression, we grouped genes into gene sets, based on their cellular and molecular functions indicated by gene ontology (GO) terms associated with the gene function descriptions. A total of 36 significant GO terms at an FDR of 10 % were identified (Supplementary Table 1). These terms are distributed in 22 biological processes (BP), 11 molecular functions (MF), and three cellular components

(CC) sub-ontologies. The relevant processes include cellular response to stimulus, cellular metabolic process, cell morphogenesis and proliferation, regulation of intracellular transport and organelle organization, and tissue development. A close look at these significant GO terms revealed several that are implicated in ASD, e.g., *neuropeptide receptor activity* (GO: 0008188) (Ramanathan et al. 2004), *neuropeptide binding* (GO: 0042923) (Baribeau and Anagnostou 2015; Lim et al. 2005), and *inhibitory synapse* (GO: 0060077) (Pettem et al. 2013; Tabuchi et al. 2007).



**Fig. 2** The breakdown of co-expression network modules in ASD. **a** Two example modules are presented as gene interaction subnetworks among non-ASD controls. Edge width is proportional to the value of Pearson's correlation coefficient (ranging 0.5–0.8). Node size is proportional to the value of  $\Delta$ SSMD for each gene. The two modules are enriched with genes whose products are closely associated with

synapse or cell junction (*top*) and genes involved in regulation of neurogenesis or neuron differentiation (*bottom*), respectively. **b** The same sets of genes in the two modules are depicted for ASD samples. The missing of edges is due to the lack of co-expression relationships between genes

### Co-expression modules that tend to be aberrantly expressed in ASD

We also used the expression data from non-ASD controls to construct the co-expression networks. We customized the parameters of WGCNA (Langfelder and Horvath 2008) (“Materials and methods”), instead of using the default values provided by the program, to construct as many as 807 network modules with relatively small size (4–110 genes). Such an adjustment of parameters was necessary because the core function `mcdcov` for MD calculation requires the size of gene sets (i.e., the size of modules) is no greater than the size of control samples ( $n = 57$ ). Otherwise, the multivariate centroid of controls would not be able to be computed. For all modules containing 57 genes or fewer, we computed SSMD and used permutation tests to assess the significance of the modules. We identified 76 significant modules that tend to be aberrantly expressed in ASD samples (Supplementary Table 2). Many genes in these modules have functions in the central nervous system. For example, module 1 is enriched with genes closely associated with synapse and cell junction while module 5 is enriched with genes involved in the regulation of neurogenesis/neuron differentiation. Figure 2a shows the co-expression relationships between genes in modules one and five among control samples. The co-expression patterns in the two modules are absent in ASD samples (Fig. 2b). It is

striking to observe such complete breakdowns of essential functional modules in ASD cases.

To quantify the module difference between ASD and control groups, we used the function `modulePreservation` in the WGCNA R package (Langfelder et al. 2011) to calculate two statistics—`medianRank` and `Zsummary`—that measure the level of connectivity preservation between modules constructed using control and ASD samples. The majority of 76 significant modules have a large `medianRank` and a close-to-zero small `Zsummary` (Supplementary Table 2), which suggest little or no module preservation across the control and ASD samples. To demonstrate that the 76 significant modules constructed using control data are robust, we obtained an independent data from brain tissues of 93 non-ASD healthy controls (GEO accession: GSE30453) (Heinzen et al. 2008) and used this new independent expression data to re-draw these significant modules. We found that, despite the difference in technical platforms (i.e., RNA sequencing vs. microarray) on which two gene expression data were generated, most co-expression relationships between genes in these modules could be recapitulated using the new independent control data—representative modules that contain ten or more genes are shown in Supplementary Figure 2. These results suggest that these modules are robust and the co-expression relationships between genes in these modules are biologically important and indispensable for healthy controls.

Next, we note that  $\Delta$ SSMD may be used as a single-gene measure to prioritize genes with desired functions. For instance, *CPLX2* is among genes with the largest value of  $\Delta$ SSMD in the module (Fig. 2a). It is likely that the sequences of *CPLX2* regulatory region are more heterogeneous among ASD samples, or the region contains variants associated with large gene expression variability more common in ASD samples. In either case,  $\Delta$ SSMD enables to prioritize gene candidates and pinpoint the genomic regions that are likely to accommodate the potential mutations responsible for the increased gene expression variability. Indeed, *CPLX2* encodes a complexin protein that binds to synaphin as part of the SNAP receptor complex and disrupts it, allowing transmitter release. *CPLX2* has been associated with schizophrenia and attention deficit hyperactivity disorder (Lee et al. 2005; Lionel et al. 2011), but not with autism yet. In future, target sequencing of the *CPLX2* regulatory region in the ASD samples may allow us to discover unknown variants associated with autism risk. Alternatively, the deregulated *CPLX2* expression might be part of the dysregulation of the entire module, which could be due to a *trans*-regulatory change (e.g., a change in a regulator of *CPLX2* and associated module). In such a case, target sequencing may be used to rule out the influence of local regulatory mutations on the *CPLX2* expression.

We also tested the correlation between  $\Delta$ SSMD and two network metrics for nodes, i.e., betweenness centrality and clustering coefficient. We previously showed that disease-causing genes have high betweenness centrality and low clustering coefficient values (Cai et al. 2010). However, for genes in these co-expression modules tested, no significant correlation was detected, which suggests  $\Delta$ SSMD captured statistical features of genes that differ from those captured by the two network metrics. Finally, we examined whether, in the same modules, genes with large  $\Delta$ SSMD tend to be expressed more differentially between ASD cases and controls. For genes in each of the 76 significant modules, we calculated *t* statistics using Student's *t* test to quantify the level of differential expression (DE) between ASD cases and controls. For each module, we then calculated the Spearman correlation coefficient ( $\rho$ ) between  $\Delta$ SSMD scores and *t* statistics of all genes. The distribution of correlation coefficients for modules is symmetrical, centered at  $\rho = 0$  with most values falling in between  $-0.5$  and  $0.5$  (Supplementary Figure 3), showing no consistent correlation between  $\Delta$ SSMD scores and DE test statistics. Thus, DE is not predictive of  $\Delta$ SSMD score or vice versa.

### Overlap between aberrantly expressed genes and ASD-implicated genes

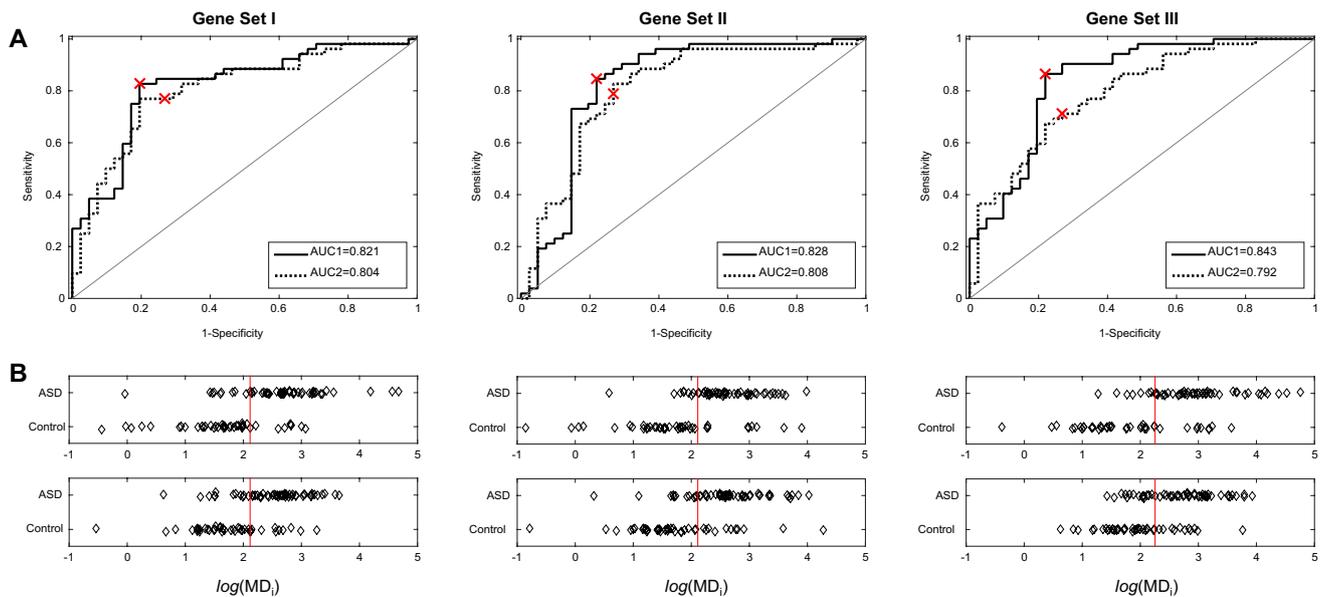
Taking all pre-defined gene sets (i.e., GSEA-defined, GO term-defined, and WGCNA modules) together, a total of

10,127 genes were under the consideration of our gene set-based analyses, and 1044 unique genes were present in the gene sets considered to be significant, for which gene expression profiles in ASD samples are over-dispersed. The overlap between these 1044 genes and 294 SFARI ASD-implicated genes (“Materials and methods”) is 36. These overlapping genes include eight of those belonging to the SFARI category of syndromic (*DHCR7*, *KCNJ10*, *MECP2*, *NF1*, *PAX6*, *SCN1A*, *TSC1* and *TSC2*), four in the category of high and strong confidence (*TBR1*, *ASXL3*, *BCL11A* and *DSCAM*), and 24 in categories of suggestive and minimal evidence. Two de novo loss-of-function mutations in *TBR1* have been previously identified in ASD patients (Neale et al. 2012; O’Roak et al. 2012a, b), along with three in *ASXL3* (De Rubeis et al. 2014; Dinwiddie et al. 2013), two in *BCL11A* (De Rubeis et al. 2014; Iossifov et al. 2012), and four in *DSCAM* (De Rubeis et al. 2014; Iossifov et al. 2014). Nevertheless, the number of overlapping genes (36) is not significantly more than expected by chance (Hypergeometric test,  $P = 0.16$ ). These results suggest that aberrant gene expression analysis, as a deviation from the *status quo*, produced the results of many novel candidate genes, which are not present in the gene list of current knowledge.

### Aberrant gene expression as biomarkers for ASD

We sought to determine whether we could classify patients as having ASD vs. controls solely based on the aberrant gene expression that is more pronounced in ASD samples. For this purpose, we obtained the gene expression data from the peripheral blood samples, including 104 ASD patients and 82 healthy controls (GEO accession: GSE18123) (Kong et al. 2012). The rationale of using this blood sample data set, rather than re-using the data set of post-mortem brains (Gupta et al. 2014), is from the position of the practical application. For diagnostic purposes, measuring gene expression in the peripheral blood makes more sense. Thus, in our analysis, a direct search for the biomarkers using the blood expression data is desired. After downloading the blood gene expression data (Kong et al. 2012), we split the data set into “training” and “test” sets, each containing data of 52 ASD and 41 control samples. With the training set data, we calculated  $MD_i$  for ASD samples against the control samples. With the test set data, we calculated  $MD_i$  for both ASD and control samples against the control samples of the training set (“Materials and methods”). That is, we calculated  $MD_i$  for all samples against the same set of controls in the training set.

Our purpose was to identify gene sets comprising several genes out of autosomal protein-coding genes expressed in the whole blood ( $n = 16,365$ ) whose aberrant gene expression could be used to distinguish ASD cases from controls (i.e.,  $MD_i$  respecting the gene sets for ASD and non-ASD



**Fig. 3** ROC curves and dot diagrams of  $MD_i$ . **a** ROC curve graphs for the three classifier gene sets tested with the training and test data sets. Corresponding AUC values for the training (AUC1) and test (AUC2) data sets are given in the inserts. *Red cross* indicates the optimal operating point of the ROC curve for the training data set. **b** *Dot*

*diagrams* for training (*top*) and test (*bottom*) sets showing the distributions of  $MD_i$  calculated with respect to the three classifier gene sets for samples in ASD and control groups. Log-transformed  $MD_i$  values are shown. The *red vertical lines* show the optimal cutoff values determined from the ROC curves tested on training data set

samples differs greatly). Here we used ROC curve analysis to evaluate the classification performance of a specific classifier gene set, so a search was conducted for gene sets with top AUC (the area under ROC curve) values based on the training set, for which the performances were assessed with the test set. We denote the AUC values with respect to the training and test data sets as AUC1 and AUC2. With randomly generated gene sets, we examined AUC2 as a function of the size of gene sets. We found no correlation between the two (Supplementary Figure 4), suggesting that random gene sets have no prediction value. The positive correlation between the size of gene sets and AUC1 (Supplementary Figure 4) is simply because that the inclusion of more variables (i.e., expression data from more genes) allows a better fit to the data. That is, the expression variation in control samples of the training set is better explained by more genes to be considered, resulting in a continuously improved AUC1. Nevertheless, the overfit of data for the training set (better AUC1) did not necessarily contribute much to the prediction for the test set (better AUC2), as suggested by the weak positive correlation between AUC1 and AUC2 (Supplementary Figure 5). Based on these preliminary results, we decided to search for gene sets containing as few as five genes to avoid the potential problem associated with overfitting of the control data. Computational time is another consideration—an exact solution for such a search for more than five genes

is a combinatorial problem requiring  $>10^{18}$  SSMD calculations, which is computationally infeasible.

To search for the five genes, we developed a greedy algorithm to search from different starting points for producing local optimal solutions (“Materials and methods”). Three gene sets, each containing five genes, were identified to generate high accuracy with balanced sensitivity and specificity values for the tests using both training and test data sets (Fig. 3). These gene sets are:  $\{FAM120A, HDC, OR13C8, PSAP, RFX8\}$ ,  $\{HBG1, MOCS3, PDGFA, SERAC1, SLFN12L\}$ , and  $\{BHMT2, CCL4L1, CD2, FAM189B, MAK\}$  (see Supplementary Table 3 for corresponding SSMD and  $\Delta$ SSMD values). All three gene sets achieved greater than 70 % sensitivity and greater than 70 % specificity in all tests (Table 2). Further analyses showed that the prediction power of the three gene sets largely remained no matter how the original data (Kong et al. 2012) were randomly split into training and test sets. Some of these classifier genes are associated with ASD, although mostly in an indirect manner. For instance, the protein product of *FAM120A* interacts with that of the ASD-implicated gene *CYFIP1* (De Rubeis et al. 2013). A rare functional mutation in *HDC*, which encodes L-histidine decarboxylase catalyzing the biosynthesis of histamine from histidine, has been associated with Tourette’s syndrome (Ercan-Sencicek et al. 2010)—a neuropsychiatric disorder potentially related to ASD (Clarke et al. 2012). The expression of *PDGFA* was found to be downregulated

**Table 2** The performances of classifiers based on gene sets I, II, III tested on the training and test data sets

Expression data set	Training set			Test set		
Number of samples (ASD + control)	93 (52 + 41)			93 (52 + 41)		
Classifier gene set	I	II	III	I	II	III
TP	43	44	45	40	41	37
TN	33	32	32	30	30	30
FP	8	9	9	11	11	11
FN	9	8	7	12	11	15
SN (%)	82.69	84.62	86.54	76.92	78.85	71.15
SP (%)	80.49	78.05	78.05	73.17	73.17	73.17
ACC (%)	81.72	81.72	82.80	75.27	76.34	72.04

True positive (TP), true negative (TN), false positive (FP), false negative (FN), sensitivity (SN), specificity (SP), accuracy (ACC) values are reported

in patients affected with the 22q11.2 deletion syndrome, which is associated with high rates of ASD in childhood (Jalbrzikowski et al. 2015).

Finally, we repeated the searching for classifier gene sets using the expression data of brain samples (Gupta et al. 2014). Three classifier gene sets were obtained: {*IFI6*, *MIDN*, *MAPK8*, *ENO2*, *GYS1*}, {*HSPH1*, *ASHIL*, *IFIT3*, *GPR3*, *PCSK2*}, and {*HNRNPK*, *GOLT1B*, *BAZ2A*, *TRABD2A*, *UNG*}, which all gave ~75 % or better prediction accuracy (Supplementary Table 4). These classifier genes show no overlap with those derived from the blood samples, but several are directly associated with ASD. For example, ASD-associated mutations have been identified in *ASHIL*, which is an SFARI category 1 (high confidence) gene (De Rubeis et al. 2014; Iossifov et al. 2014; Tammimies et al. 2015; Willsey et al. 2013) and in *MIDN*, which is involved in neurogenesis and neuronal migration (Butler et al. 2015).

## Discussion

ASD is a complex disease involving multiple genetic alterations that result in modifications of many cellular processes. Maladaptive patterns of ASD lead to significantly high gene expression variability among affected individuals. Unitary models of autism brain dysfunction have not adequately addressed conflicting evidence, and efforts to find a single unifying brain dysfunction have led the field away from research to explore individual variation and micro-subgroups. Therefore, it has been suggested that researchers must explore individual variation in brain measures within autism (Geschwind 2008; Waterhouse and Gillberg 2014). Previous studies, with few exceptions (Garbett et al. 2008; Voineagu et al. 2011), have rarely addressed the issue of increased gene expression variability associated with autism. Noticeably, Voineagu et al. (2011) pointed out:

“Autistic subjects display significant phenotypic variability which could be due to an intricate interplay of genetic and environmental factors. Thus, we hypothesized that this phenotypic diversity is due to subject-to-subject variability in gene expression.” Nevertheless, the status quo pertaining to gene expression specific to ASD patients is based on the detection of differential gene expression, i.e., the gene expression differences between mathematical expectation (i.e., mean) of ASD and control samples. The major assumption underlying differential expression analysis is: ASD cases have the same or similar gene expression change phenotypes, which makes them as a separate cohort have significantly higher or lower expression than the controls. However, this assumption contradicts the fact that ASD has highly heterogeneous genetic causes, and excludes empirical evidence gathered about uncommon molecular changes causing ASD (Neale et al. 2012; Pinto et al. 2014; Sanders et al. 2012; Sebat et al. 2007).

## Dispersion-specific measure of gene expression for autism

Our overall strategy for this study was based on the quantitative measures of the departure of multigene expression dispersion between individuals. The profound heterogeneity in ASD underscores the importance of leveraging measures of dispersion to capture the specific tendency. Gene expression dispersion has been found associated with gene function and disease or physiological status of individuals (Ecker et al. 2015; Li et al. 2010; Mar et al. 2011; Somel et al. 2006). Discrepancies in gene expression should not only be characterized by the mean but also by other statistics of interest, such as dispersion parameters. Using the proven multivariate approach (Zeng et al. 2015), we have further developed MD-based aberrant gene expression analysis and applied it to ASD. The statistical signal captured is the tendency of being more dispersed in multigene

expression among ASD than control samples. We have shown that our variability-centered method can recapitulate signals from many genes known to be implicated in ASD. Our method does not depend on the prior knowledge about gene function or the identification of mutations in genes. Thus, it is a tool for discovering and identifying genes previously unknown to be involved in ASD progression.

### Aberrant gene expression in co-expression network modules

We have shown that, when applied to the co-expression network, SSMD can reveal the effects of perturbing genetic networks. SSMD analysis informs us about how ASD distorts expression patterns of biological systems. Disturbed ASD genetic networks have been noticed previously (Hormozdiari et al. 2015; Li et al. 2014; Parikshak et al. 2013; Pramparo et al. 2015; Willsey et al. 2013). However, most existing network analyses were not designed for directly measuring the level of dysregulation. Instead, information about known ASD genes, e.g., in (Li et al. 2014; Willsey et al. 2013), or differently expressed genes, e.g., in (Pramparo et al. 2015), was used to prioritize the modules, which would not allow modules contain unknown ASD genes to be prioritized for subsequent analyses. In contrast, our approach allows for a straightforward screening of perturbed network modules and provides the raw material for the identification of genetic regulatory mechanisms involved in the variability of gene transcription.

### Toward the genetic basis of aberrant gene expression

Our results have provided unique entry points to investigate further on the genetic basis of aberrant gene expression (e.g., increased gene expression variability) in ASD. When genotype or sequence information, along with their gene expression information, becomes available for ASD samples, it would be possible to assess the influences of the aggregation of rare mutations, CNVs, as well as common genetic variants on aberrant patterns of gene expression in ASD. In line with this view, the latest genome sequencing effort for autism-affected families showed that disruptive *de novo*/private mutations and CNVs are significantly enriched in regulatory regions of ASD-related genes in ASD probands (Turner et al. 2016). Furthermore, we have shown previously that certain common genetic variants, in addition to rare variants, cause the increase of gene expression variability among individuals (Hulse and Cai 2013). These common variants influence the variability of gene expression through the action of either epistasis or direct destabilization (Wang et al. 2014). By taking both rare and common variants into account, it would be possible to superimpose their impact onto a gene expression variability

network to predict which parts of the network are more vulnerable to the perturbation from genetic factors such as ASD-related disruptive mutations.

### Aberrant gene expression as biomarkers

Recent years have seen an intensive search for biological markers for ASD. Although a wide range of ASD biomarkers has been proposed, as of yet none has been validated for clinical use (Walsh et al. 2011). Therefore, there is a critical need for valid biological markers for ASD. Based on the results of aberrant gene expression analysis shown here, gene sets with just a few selected genes can be used as novel biomarkers. Application of our gene expression candidate biomarkers will allow for higher sensitivity and specificity in a diagnostic screen for ASD. We anticipate that if our gene expression biomarkers are expanded to use the blood gene expression data derived from other platforms (such as different types of microarrays, RNA sequencing, and qPCR), they will offer a significant advancement in developing a clinical blood test. The success of such gene expression biomarkers will assist in early and objective diagnosis for ASD.

### Caveats and future directions

Voineagu et al. (2011) showed that the heterogeneity in gene expression between different brain regions of the same individual might introduce another level of gene expression variability. Due to the limitation of tissue samples available for this study, such an effect was not explicitly captured by our aberrant gene expression analysis. Also, brain regions themselves are highly heterogeneous because of the mixtures of cell types. Aberrant gene expression patterns might in part indicate different relative proportions of cell types in a sample. With the advent of the single-cell-based technologies (Dey et al. 2015), this level of gene expression heterogeneity may be measured. Thus, the problem of heterogeneity of cell types in tissue samples as an important source of variability may be addressed in future studies.

### Conclusions

We have developed a novel, variability-centric gene expression analysis, and applied the method to ASD. This advance showcases the value of development and refinement of systems genomics tools in studying human complex diseases. The aberrantly expressed genes identified in this study will facilitate the identification of ASD-predisposing variation, which may eventually reveal the causes of ASD and enable earlier and more targeted methods for diagnosis and intervention.

**Acknowledgments** We thank Shannon Ellis and Dan Arking for sharing the data, Oliver Stegle and Tuuli Lappalainen for helping with data normalization, and Steve Horvath for the co-expression network analysis. We thank Rae L. Russell for proofreading and editing this paper. We acknowledge the Texas A&M Institute for Genome Sciences and Society (TIGSS) for providing computing resources and system administration support. This work was supported by the fund of China Scholarship Council to JG, and the National Natural Science Foundation of China (No. 61573296), the Specialized Research Fund for the Doctoral Program of Higher Education of China (No. 20130121130004), the Fundamental Research Funds for the Central Universities in China (Xiamen University: Nos. 201412G009, 201510384106) to GJ.

#### Compliance with ethical standards

**Competing interests** The authors declare that they have no competing interests.

## References

- Abrahams BS, Geschwind DH (2008) Advances in autism genetics: on the threshold of a new neurobiology. *Nat Rev Genet* 9:341–355. doi:[10.1038/nrg2346](https://doi.org/10.1038/nrg2346)
- Baribeau DA, Anagnostou E (2015) Oxytocin and vasopressin: linking pituitary neuropeptides and their receptors to social neurocircuits. *Front Neurosci* 9:335. doi:[10.3389/fnins.2015.00335](https://doi.org/10.3389/fnins.2015.00335)
- Belmonte MK, Cook EH Jr, Anderson GM, Rubenstein JL, Greenough WT, Beckel-Mitchener A, Courchesne E, Boulanger LM, Powell SB, Levitt PR, Perry EK, Jiang YH, DeLorey TM, Tierney E (2004) Autism as a disorder of neural information processing: directions for research and targets for therapy. *Mol Psychiatry* 9:646–663. doi:[10.1038/sj.mp.4001499](https://doi.org/10.1038/sj.mp.4001499)
- Benfenati F, Valtorta F, Bahler M, Greengard P (1989) Synapsin I, a neuron-specific phosphoprotein interacting with small synaptic vesicles and F-actin. *Cell Biol Int Rep* 13:1007–1021
- Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate—a practical and powerful approach to multiple testing. *J R Stat Soc Ser B Methodol* 57:289–300. doi:[10.2307/2346101](https://doi.org/10.2307/2346101)
- Butler MG, Rafi SK, Hossain W, Stephan DA, Manzardo AM (2015) Whole exome sequencing in females with autism implicates novel and candidate genes. *Int J Mol Sci* 16:1312–1335. doi:[10.3390/ijms16011312](https://doi.org/10.3390/ijms16011312)
- Cai JJ, Borenstein E, Petrov DA (2010) Broker genes in human disease. *Genome Biol Evol* 2:815–825. doi:[10.1093/gbe/evq064](https://doi.org/10.1093/gbe/evq064)
- Cesca F, Baldelli P, Valtorta F, Benfenati F (2010) The synapsins: key actors of synapse function and plasticity. *Prog Neurobiol* 91:313–348. doi:[10.1016/j.pneurobio.2010.04.006](https://doi.org/10.1016/j.pneurobio.2010.04.006)
- Chow ML, Pramparo T, Winn ME, Barnes CC, Li HR, Weiss L, Fan JB, Murray S, April C, Belinson H, Fu XD, Wynshaw-Boris A, Schork NJ, Courchesne E (2012) Age-dependent brain gene expression and copy number anomalies in autism suggest distinct pathological processes at young versus mature ages. *PLoS Genet* 8:e1002592. doi:[10.1371/journal.pgen.1002592](https://doi.org/10.1371/journal.pgen.1002592)
- Clarke RA, Lee S, Eapen V (2012) Pathogenetic model for Tourette syndrome delineates overlap with related neurodevelopmental disorders including autism. *Transl Psychiatry* 2:e158. doi:[10.1038/tp.2012.75](https://doi.org/10.1038/tp.2012.75)
- Cookson W, Liang L, Abecasis G, Moffatt M, Lathrop M (2009) Mapping complex disease traits with global gene expression. *Nat Rev Genet* 10:184–194. doi:[10.1038/nrg2537](https://doi.org/10.1038/nrg2537)
- Corradi A, Fadda M, Piton A, Patry L, Marte A, Rossi P, Cadieux-Dion M, Gauthier J, Lapointe L, Mottron L, Valtorta F, Rouleau GA, Fassio A, Benfenati F, Cossette P (2014) SYN2 is an autism predisposing gene: loss-of-function mutations alter synaptic vesicle cycling and axon outgrowth. *Hum Mol Genet* 23:90–103. doi:[10.1093/hmg/ddt401](https://doi.org/10.1093/hmg/ddt401)
- Croft D, Mundo AF, Haw R, Milacic M, Weiser J, Wu G, Caudy M, Garapati P, Gillespie M, Kamdar MR, Jassal B, Jupe S, Matthews L, May B, Palatnik S, Rothfels K, Shamovsky V, Song H, Williams M, Birney E, Hermjakob H, Stein L, D'Eustachio P (2014) The Reactome pathway knowledgebase. *Nucleic Acids Res* 42:D472–D477. doi:[10.1093/nar/gkt1102](https://doi.org/10.1093/nar/gkt1102)
- De Rubeis S, Pasciuto E, Li KW, Fernandez E, Di Marino D, Buzzi A, Ostroff LE, Klann E, Zwartkruis FJ, Komiyama NH, Grant SG, Poujol C, Choquet D, Achsel T, Posthuma D, Smit AB, Bagni C (2013) CYFIP1 coordinates mRNA translation and cytoskeleton remodeling to ensure proper dendritic spine formation. *Neuron* 79:1169–1182. doi:[10.1016/j.neuron.2013.06.039](https://doi.org/10.1016/j.neuron.2013.06.039)
- De Rubeis S, He X, Goldberg AP, Poultney CS, Samocha K, Cicek AE, Kou Y, Liu L, Fromer M, Walker S, Singh T, Klei L, Kosmicki J, Shih-Chen F, Aleksic B, Biscaldi M, Bolton PF, Brownfeld JM, Cai J, Campbell NG, Carracedo A, Chahrour MH, Chiochetti AG, Coon H, Crawford EL, Curran SR, Dawson G, Duketis E, Fernandez BA, Gallagher L, Geller E, Guter SJ, Hill RS, Ionita-Laza J, Jimenez Gonzalez P, Kilpinen H, Klauck SM, Kolevzon A, Lee I, Lei J, Lehtimaki T, Lin CF, Ma'ayan A, Marshall CR, McInnes AL, Neale B, Owen MJ, Ozaki N, Parelada M, Parr JR, Purcell S, Puura K, Rajagopalan D, Rehnstrom K, Reichenberg A, Sabo A, Sachse M, Sanders SJ, Schaffer C, Schulte-Ruther M, Skuse D, Stevens C, Szatmari P, Tammimies K, Valladares O, Voran A, Li-San W, Weiss LA, Willsey AJ, Yu TW, Yuen RK, DDD Study, Homozygosity Mapping Collaborative for Autism, UK10K Consortium, Cook EH, Freitag CM, Gill M, Hultman CM, Lehner T, Palotie A, Schellenberg GD, Sklar P, State MW, Sutcliffe JS, Walsh CA, Scherer SW, Zwick ME, Barrett JC, Cutler DJ, Roeder K, Devlin B, Daly MJ, Buxbaum JD (2014) Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* 515:209–215. doi:[10.1038/nature13772](https://doi.org/10.1038/nature13772)
- Devlin B, Scherer SW (2012) Genetic architecture in autism spectrum disorder. *Curr Opin Genet Dev* 22:229–237. doi:[10.1016/j.gde.2012.03.002](https://doi.org/10.1016/j.gde.2012.03.002)
- Dey SS, Foley JE, Limsirichai P, Schaffer DV, Arkin AP (2015) Orthogonal control of expression mean and variance by epigenetic features at different genomic loci. *Mol Syst Biol* 11:806. doi:[10.15252/msb.20145704](https://doi.org/10.15252/msb.20145704)
- Dinwiddie DL, Soden SE, Saunders CJ, Miller NA, Farrow EG, Smith LD, Kingsmore SF (2013) De novo frameshift mutation in ASXL3 in a patient with global developmental delay, microcephaly, and craniofacial anomalies. *BMC Med Genomics* 6:32. doi:[10.1186/1755-8794-6-32](https://doi.org/10.1186/1755-8794-6-32)
- Ecker S, Pancaldi V, Rico D, Valencia A (2015) Higher gene expression variability in the more aggressive subtype of chronic lymphocytic leukemia. *Genome Med* 7:8. doi:[10.1186/s13073-014-0125-z](https://doi.org/10.1186/s13073-014-0125-z)
- Elsabbagh M, Johnson MH (2010) Getting answers from babies about autism. *Trends Cogn Sci* 14:81–87. doi:[10.1016/j.tics.2009.12.005](https://doi.org/10.1016/j.tics.2009.12.005)
- Ercan-Sencicek AG, Stillman AA, Ghosh AK, Bilguvar K, O'Roak BJ, Mason CE, Abbott T, Gupta A, King RA, Pauls DL, Tischfield JA, Heiman GA, Singer HS, Gilbert DL, Hoekstra PJ, Morgan TM, Loring E, Yasuno K, Fernandez T, Sanders S, Louvi A, Cho JH, Mane S, Colangelo CM, Biederer T, Lifton RP, Gunel M, State MW (2010) L-Histidine decarboxylase and Tourette's syndrome. *N Engl J Med* 362:1901–1908. doi:[10.1056/NEJMoa0907006](https://doi.org/10.1056/NEJMoa0907006)
- Fatemi SH, Folsom TD, Reutiman TJ, Sidwell RW (2008) Viral regulation of aquaporin 4, connexin 43, microcephalin and nucleolin. *Schizophr Res* 98:163–177. doi:[10.1016/j.schres.2007.09.031](https://doi.org/10.1016/j.schres.2007.09.031)

- Flint J, Timpson N, Munafò M (2014) Assessing the utility of intermediate phenotypes for genetic mapping of psychiatric disease. *Trends Neurosci* 37:733–741. doi:[10.1016/j.tins.2014.08.007](https://doi.org/10.1016/j.tins.2014.08.007)
- Frye RE, Huffman LC, Elliott GR (2010) Tetrahydrobiopterin as a novel therapeutic intervention for autism. *Neurotherapeutics* 7:241–249. doi:[10.1016/j.nurt.2010.05.004](https://doi.org/10.1016/j.nurt.2010.05.004)
- Garbett K, Ebert PJ, Mitchell A, Lintas C, Manzi B, Mirnics K, Persico AM (2008) Immune transcriptome alterations in the temporal cortex of subjects with autism. *Neurobiol Dis* 30:303–311. doi:[10.1016/j.nbd.2008.01.012](https://doi.org/10.1016/j.nbd.2008.01.012)
- Gaugler T, Klei L, Sanders SJ, Bodea CA, Goldberg AP, Lee AB, Mahajan M, Manaa D, Pawitan Y, Reichert J, Ripke S, Sandin S, Sklar P, Svantesson O, Reichenberg A, Hultman CM, Devlin B, Roeder K, Buxbaum JD (2014) Most genetic risk for autism resides with common variation. *Nat Genet* 46:881–885. doi:[10.1038/ng.3039](https://doi.org/10.1038/ng.3039)
- Geschwind DH (2008) Autism: many genes, common pathways? *Cell* 135:391–395. doi:[10.1016/j.cell.2008.10.016](https://doi.org/10.1016/j.cell.2008.10.016)
- Geschwind DH (2011) Genetics of autism spectrum disorders. *Trends Cogn Sci* 15:409–416. doi:[10.1016/j.tics.2011.07.003](https://doi.org/10.1016/j.tics.2011.07.003)
- Geschwind DH, State MW (2015) Gene hunting in autism spectrum disorder: on the path to precision medicine. *Lancet Neurol* 14:1109–1120. doi:[10.1016/S1474-4422\(15\)00044-7](https://doi.org/10.1016/S1474-4422(15)00044-7)
- Glessner JT, Wang K, Cai G, Korvatska O, Kim CE, Wood S, Zhang H, Estes A, Brune CW, Bradfield JP, Imielinski M, Frackelton EC, Reichert J, Crawford EL, Munson J, Sleiman PM, Chiavacci R, Annaiah K, Thomas K, Hou C, Glaberson W, Flory J, Otieno F, Garris M, Soorya L, Klei L, Piven J, Meyer KJ, Anagnostou E, Sakurai T, Game RM, Rudd DS, Zurawiecki D, McDougle CJ, Davis LK, Miller J, Posey DJ, Michaels S, Kolevzon A, Silverman JM, Bernier R, Levy SE, Schultz RT, Dawson G, Owley T, McMahon WM, Wassink TH, Sweeney JA, Nurnberger JI, Coon H, Sutcliffe JS, Minshew NJ, Grant SF, Bucan M, Cook EH, Buxbaum JD, Devlin B, Schellenberg GD, Hakonarson H (2009) Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. *Nature* 459:569–573. doi:[10.1038/nature07953](https://doi.org/10.1038/nature07953)
- Guglielmi L, Servettini I, Caramia M, Catacuzzeno L, Franciolini F, D'Adamo MC, Pessia M (2015) Update on the implication of potassium channels in autism: K(+) channelautism spectrum disorder. *Front Cell Neurosci* 9:34. doi:[10.3389/fncel.2015.00034](https://doi.org/10.3389/fncel.2015.00034)
- Gupta S, Ellis SE, Ashar FN, Moes A, Bader JS, Zhan J, West AB, Arking DE (2014) Transcriptome analysis reveals dysregulation of innate immune response genes and neuronal activity-dependent genes in autism. *Nat Commun* 5:5748. doi:[10.1038/ncomms6748](https://doi.org/10.1038/ncomms6748)
- Hansen KD, Irizarry RA, Wu Z (2012) Removing technical variability in RNA-seq data using conditional quantile normalization. *Bio-statistics* 13:204–216. doi:[10.1093/biostatistics/kxr054](https://doi.org/10.1093/biostatistics/kxr054)
- Heinzen EL, Ge D, Cronin KD, Maia JM, Shianna KV, Gabriel WN, Welsh-Bohmer KA, Hulette CM, Denny TN, Goldstein DB (2008) Tissue-specific genetic control of splicing: implications for the study of complex traits. *PLoS Biol* 6:e1. doi:[10.1371/journal.pbio.1000001](https://doi.org/10.1371/journal.pbio.1000001)
- Hormozdiari F, Penn O, Borenstein E, Eichler EE (2015) The discovery of integrated gene networks for autism and related disorders. *Genome Res* 25:142–154. doi:[10.1101/gr.178855.114](https://doi.org/10.1101/gr.178855.114)
- Hulse AM, Cai JJ (2013) Genetic variants contribute to gene expression variability in humans. *Genetics* 193:95–108. doi:[10.1534/genetics.112.146779](https://doi.org/10.1534/genetics.112.146779)
- Iossifov I, Ronemus M, Levy D, Wang Z, Hakker I, Rosenbaum J, Yamrom B, Lee YH, Narzisi G, Leotta A, Kendall J, Grabowska E, Ma B, Marks S, Rodgers L, Stepansky A, Troge J, Andrews P, Bekrisky M, Pradhan K, Ghiban E, Kramer M, Parla J, Demeter R, Fulton LL, Fulton RS, Magrini VJ, Ye K, Darnell JC, Darnell RB, Mardis ER, Wilson RK, Schatz MC, McCombie WR, Wigler M (2012) De novo gene disruptions in children on the autistic spectrum. *Neuron* 74:285–299. doi:[10.1016/j.neuron.2012.04.009](https://doi.org/10.1016/j.neuron.2012.04.009)
- Iossifov I, O'Roak BJ, Sanders SJ, Ronemus M, Krumm N, Levy D, Stessman HA, Witherspoon KT, Vives L, Patterson KE, Smith JD, Paepfer B, Nickerson DA, Dea J, Dong S, Gonzalez LE, Mandell JD, Mane SM, Murtha MT, Sullivan CA, Walker MF, Waqar Z, Wei L, Willsey AJ, Yamrom B, Lee YH, Grabowska E, Dalkic E, Wang Z, Marks S, Andrews P, Leotta A, Kendall J, Hakker I, Rosenbaum J, Ma B, Rodgers L, Troge J, Narzisi G, Yoon S, Schatz MC, Ye K, McCombie WR, Shendure J, Eichler EE, State MW, Wigler M (2014) The contribution of de novo coding mutations to autism spectrum disorder. *Nature* 515:216–221. doi:[10.1038/nature13908](https://doi.org/10.1038/nature13908)
- Jalbrzikowski M, Lazaro MT, Gao F, Huang A, Chow C, Geschwind DH, Coppola G, Bearden CE (2015) Transcriptome profiling of peripheral blood in 22q11.2 deletion syndrome reveals functional pathways related to psychosis and autism spectrum disorder. *PLoS One* 10:e0132542. doi:[10.1371/journal.pone.0132542](https://doi.org/10.1371/journal.pone.0132542)
- Klaiman C, Huffman L, Masaki L, Elliott GR (2013) Tetrahydrobiopterin as a treatment for autism spectrum disorders: a double-blind, placebo-controlled trial. *J Child Adolesc Psychopharmacol* 23:320–328. doi:[10.1089/cap.2012.0127](https://doi.org/10.1089/cap.2012.0127)
- Klei L, Sanders SJ, Murtha MT, Hus V, Lowe JK, Willsey AJ, Moreno-De-Luca D, Yu TW, Fombonne E, Geschwind D, Grice DE, Ledbetter DH, Lord C, Mane SM, Martin CL, Martin DM, Morrow EM, Walsh CA, Melhem NM, Chaste P, Sutcliffe JS, State MW, Cook EH Jr, Roeder K, Devlin B (2012) Common genetic variants, acting additively, are a major source of risk for autism. *Mol Autism* 3:9. doi:[10.1186/2040-2392-3-9](https://doi.org/10.1186/2040-2392-3-9)
- Kong SW, Collins CD, Shimizu-Motohashi Y, Holm IA, Campbell MG, Lee IH, Brewster SJ, Hanson E, Harris HK, Lowe KR, Saada A, Mora A, Madison K, Hundley R, Egan J, McCarthy J, Eran A, Galdzicki M, Rappaport L, Kunkel LM, Kohane IS (2012) Characteristics and predictive value of blood transcriptome signature in males with autism spectrum disorders. *PLoS One* 7:e49475. doi:[10.1371/journal.pone.0049475](https://doi.org/10.1371/journal.pone.0049475)
- Kongatik K, Wang G, Yang E, Cai JJ (2013) SBEToolbox: a Matlab toolbox for biological network analysis. *Evol Bioinform Online* 9:355–362. doi:[10.4137/EBO.S12012](https://doi.org/10.4137/EBO.S12012)
- Langfelder P, Horvath S (2008) WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9:559. doi:[10.1186/1471-2105-9-559](https://doi.org/10.1186/1471-2105-9-559)
- Langfelder P, Luo R, Oldham MC, Horvath S (2011) Is my network module preserved and reproducible? *PLoS Comput Biol* 7:e1001057. doi:[10.1371/journal.pcbi.1001057](https://doi.org/10.1371/journal.pcbi.1001057)
- Lazaro MT, Golshani P (2015) The utility of rodent models of autism spectrum disorders. *Curr Opin Neurol* 28:103–109. doi:[10.1097/WCO.0000000000000183](https://doi.org/10.1097/WCO.0000000000000183)
- Lee HJ, Song JY, Kim JW, Jin SY, Hong MS, Park JK, Chung JH, Shibata H, Fukumaki Y (2005) Association study of polymorphisms in synaptic vesicle-associated genes, SYN2 and CPLX2, with schizophrenia. *Behav Brain Funct* 1:15. doi:[10.1186/1744-9081-1-15](https://doi.org/10.1186/1744-9081-1-15)
- Lee H, Lin MC, Kornblum HI, Papazian DM, Nelson SF (2014) Exome sequencing identifies de novo gain of function missense mutation in KCND2 in identical twins with autism and seizures that slows potassium channel inactivation. *Hum Mol Genet* 23:3481–3489. doi:[10.1093/hmg/ddu056](https://doi.org/10.1093/hmg/ddu056)
- Li J, Liu Y, Kim T, Min R, Zhang Z (2010) Gene expression variability within and between human populations and implications toward disease susceptibility. *PLoS Comput Biol* 6:e1000910. doi:[10.1371/journal.pcbi.1000910](https://doi.org/10.1371/journal.pcbi.1000910)
- Li J, Shi M, Ma Z, Zhao S, Euskirchen G, Ziskin J, Urban A, Hallmayer J, Snyder M (2014) Integrated systems analysis reveals a

- molecular network underlying autism spectrum disorders. *Mol Syst Biol* 10:774. doi:[10.15252/msb.20145487](https://doi.org/10.15252/msb.20145487)
- Liberzon A, Subramanian A, Pinchback R, Thorvaldsdottir H, Tamayo P, Mesirov JP (2011) Molecular signatures database (MSigDB) 3.0. *Bioinformatics* 27:1739–1740. doi:[10.1093/bioinformatics/btr260](https://doi.org/10.1093/bioinformatics/btr260)
- Lim MM, Bielsky IF, Young LJ (2005) Neuropeptides and the social brain: potential rodent models of autism. *Int J Dev Neurosci* 23:235–243. doi:[10.1016/j.jdevneu.2004.05.006](https://doi.org/10.1016/j.jdevneu.2004.05.006)
- Lionel AC, Crosbie J, Barbosa N, Goodale T, Thiruvahindrapuram B, Rickaby J, Gazzellone M, Carson AR, Howe JL, Wang Z, Wei J, Stewart AF, Roberts R, McPherson R, Fiebig A, Franke A, Schreiber S, Zwaigenbaum L, Fernandez BA, Roberts W, Arnold PD, Szatmari P, Marshall CR, Schachar R, Scherer SW (2011) Rare copy number variation discovery and cross-disorder comparisons identify risk genes for ADHD. *Sci Transl Med* 3:95ra75. doi:[10.1126/scitranslmed.3002464](https://doi.org/10.1126/scitranslmed.3002464)
- Mahalanobis PC (1936) On the generalised distance in statistics. *Proc Natl Inst Sci India* 2:49–55. doi:[citeulike-article-id:4155812](https://doi.org/10.1007/bf02854112)
- Mar JC, Matigian NA, Mackay-Sim A, Mellick GD, Sue CM, Silburn PA, McGrath JJ, Quackenbush J, Wells CA (2011) Variance of gene expression identifies altered network constraints in neurological disease. *PLoS Genet* 7:e1002207. doi:[10.1371/journal.pgen.1002207](https://doi.org/10.1371/journal.pgen.1002207)
- Neale BM, Kou Y, Liu L, Ma'ayan A, Samocha KE, Sabo A, Lin CF, Stevens C, Wang LS, Makarov V, Polak P, Yoon S, Maguire J, Crawford EL, Campbell NG, Geller ET, Valladares O, Schafer C, Liu H, Zhao T, Cai G, Lihm J, Dannenfelser R, Jabado O, Peralta Z, Nagaswamy U, Muzny D, Reid JG, Newsham I, Wu Y, Lewis L, Han Y, Voight BF, Lim E, Rossin E, Kirby A, Flanick J, Fromer M, Shakir K, Fennell T, Garimella K, Banks E, Poplin R, Gabriel S, DePristo M, Wimbish JR, Boone BE, Levy SE, Betancur C, Sunyaev S, Boerwinkle E, Buxbaum JD, Cook EH Jr, Devlin B, Gibbs RA, Roeder K, Schellenberg GD, Sutcliffe JS, Daly MJ (2012) Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature* 485:242–245. doi:[10.1038/nature11011](https://doi.org/10.1038/nature11011)
- O'Roak BJ, Vives L, Fu W, Egerton JD, Stanaway IB, Phelps IG, Carvill G, Kumar A, Lee C, Ankenman K, Munson J, Hiatt JB, Turner EH, Levy R, O'Day DR, Krumm N, Coe BP, Martin BK, Borenstein E, Nickerson DA, Mefford HC, Doherty D, Akey JM, Bernier R, Eichler EE, Shendure J (2012a) Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. *Science* 338:1619–1622. doi:[10.1126/science.1227764](https://doi.org/10.1126/science.1227764)
- O'Roak BJ, Vives L, Girirajan S, Karakoc E, Krumm N, Coe BP, Levy R, Ko A, Lee C, Smith JD, Turner EH, Stanaway IB, Vernot B, Malig M, Baker C, Reilly B, Akey JM, Borenstein E, Rieder MJ, Nickerson DA, Bernier R, Shendure J, Eichler EE (2012b) Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature* 485:246–250. doi:[10.1038/nature10989](https://doi.org/10.1038/nature10989)
- Parikshak NN, Luo R, Zhang A, Won H, Lowe JK, Chandran V, Horvath S, Geschwind DH (2013) Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. *Cell* 155:1008–1021. doi:[10.1016/j.cell.2013.10.031](https://doi.org/10.1016/j.cell.2013.10.031)
- Persico AM, Bourgeron T (2006) Searching for ways out of the autism maze: genetic, epigenetic and environmental clues. *Trends Neurosci* 29:349–358. doi:[10.1016/j.tins.2006.05.010](https://doi.org/10.1016/j.tins.2006.05.010)
- Pettem KL, Yokomaku D, Takahashi H, Ge Y, Craig AM (2013) Interaction between autism-linked MDGAs and neuroligins suppresses inhibitory synapse development. *J Cell Biol* 200:321–336. doi:[10.1083/jcb.201206028](https://doi.org/10.1083/jcb.201206028)
- Pinto D, Delaby E, Merico D, Barbosa M, Merikangas A, Klei L, Thiruvahindrapuram B, Xu X, Ziman R, Wang Z, Vorstman JA, Thompson A, Regan R, Pilorge M, Pellecchia G, Pagnamenta AT, Oliveira B, Marshall CR, Magalhaes TR, Lowe JK, Howe JL, Griswold AJ, Gilbert J, Duketis E, Dombroski BA, De Jonge MV, Cuccaro M, Crawford EL, Correia CT, Conroy J, Conceicao IC, Chiochetti AG, Casey JP, Cai G, Cabrol C, Bolshakova N, Bacchelli E, Anney R, Gallinger S, Cotterchio M, Casey G, Zwaigenbaum L, Wittmeyer K, Wing K, Wallace S, van Engeland H, Tryfon A, Thomson S, Soorya L, Roge B, Roberts W, Poustka F, Moug S, Minshew N, McInnes LA, McGrew SG, Lord C, Leboyer M, Le Couteur AS, Kolevzon A, Jimenez Gonzalez P, Jacob S, Holt R, Guter S, Green J, Green A, Gillberg C, Fernandez BA, Duque F, Delorme R, Dawson G, Chaste P, Cafe C, Brennan S, Bourgeron T, Bolton PF, Bolte S, Bernier R, Baird G, Bailey AJ, Anagnostou E, Almeida J, Wijsman EM, Vieland VJ, Vicente AM, Schellenberg GD, Pericak-Vance M, Paterson AD, Parr JR, Oliveira G, Nurnberger JI, Monaco AP, Maestrini E, Klauck SM, Hakonarson H, Haines JL, Geschwind DH, Freitag CM, Folstein SE, Ennis S et al (2014) Convergence of genes and cellular pathways dysregulated in autism spectrum disorders. *Am J Hum Genet* 94:677–694. doi:[10.1016/j.ajhg.2014.03.018](https://doi.org/10.1016/j.ajhg.2014.03.018)
- Pramparo T, Pierce K, Lombardo MV, Carter Barnes C, Marinero S, Ahrens-Barbeau C, Murray SS, Lopez L, Xu R, Courchesne E (2015) Prediction of autism by translation and immune/inflammation coexpressed genes in toddlers from pediatric community practices. *JAMA Psychiatry* 72:386–394. doi:[10.1001/jamapsychiatry.2014.3008](https://doi.org/10.1001/jamapsychiatry.2014.3008)
- Ramanathan S, Woodroffe A, Flodman PL, Mays LZ, Hanouni M, Modahl CB, Steinberg-Epstein R, Bocian ME, Spence MA, Smith M (2004) A case of autism with an interstitial deletion on 4q leading to hemizyosity for genes encoding for glutamine and glycine neurotransmitter receptor sub-units (AMPA 2, GLRA3, GLRB) and neuropeptide receptors NPY1R, NPY5R. *BMC Med Genet* 5:10. doi:[10.1186/1471-2350-5-10](https://doi.org/10.1186/1471-2350-5-10)
- Rousseeuw PJ, Van Driessen K (1999) A fast algorithm for the minimum covariance determinant estimator. *Technometrics* 41:212–223. doi:[10.2307/1270566](https://doi.org/10.2307/1270566)
- Sanders SJ, Murtha MT, Gupta AR, Murdoch JD, Raubeson MJ, Willis AJ, Ercan-Sencicek AG, DiLullo NM, Parikshak NN, Stein JL, Walker MF, Ober GT, Teran NA, Song Y, El-Fishawy P, Murtha RC, Choi M, Overton JD, Bjornson RD, Carriero NJ, Meyer KA, Bilguvar K, Mane SM, Sestan N, Lifton RP, Gunel M, Roeder K, Geschwind DH, Devlin B, State MW (2012) De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* 485:237–241. doi:[10.1038/nature10945](https://doi.org/10.1038/nature10945)
- Sandin S, Lichtenstein P, Kuja-Halkola R, Larsson H, Hultman CM, Reichenberg A (2014) The familial risk of autism. *JAMA* 311:1770–1777. doi:[10.1001/jama.2014.4144](https://doi.org/10.1001/jama.2014.4144)
- Sawicka K, Zukin RS (2012) Dysregulation of mTOR signaling in neuropsychiatric disorders: therapeutic implications. *Neuropsychopharmacology* 37:305–306. doi:[10.1038/npp.2011.210](https://doi.org/10.1038/npp.2011.210)
- Schaaf CP, Sabo A, Sakai Y, Crosby J, Muzny D, Hawes A, Lewis L, Akbar H, Varghese R, Boerwinkle E, Gibbs RA, Zoghbi HY (2011) Oligogenic heterozygosity in individuals with high-functioning autism spectrum disorders. *Hum Mol Genet* 20:3366–3375. doi:[10.1093/hmg/ddr243](https://doi.org/10.1093/hmg/ddr243)
- Schubert D, Martens GJ, Kolk SM (2015) Molecular underpinnings of prefrontal cortex development in rodents provide insights into the etiology of neurodevelopmental disorders. *Mol Psychiatry* 20:795–809. doi:[10.1038/mp.2014.147](https://doi.org/10.1038/mp.2014.147)
- Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-Martin C, Walsh T, Yamrom B, Yoon S, Krasnitz A, Kendall J, Leotta A, Pai D, Zhang R, Lee YH, Hicks J, Spence SJ, Lee AT, Puura K, Lehtimäki T, Ledbetter D, Gregersen PK, Bregman J, Sutcliffe JS, Jobanputra V, Chung W, Warburton D, King MC, Skuse D, Geschwind DH, Gilliam TC, Ye K, Wigler M (2007) Strong

- association of de novo copy number mutations with autism. *Science* 316:445–449. doi:[10.1126/science.1138659](https://doi.org/10.1126/science.1138659)
- Smedley D, Haider S, Durinck S, Pandini L, Provero P, Allen J, Arnaiz O, Awedh MH, Baldock R, Barbiera G, Bardou P, Beck T, Blake A, Bonierbale M, Brookes AJ, Bucci G, Buetti I, Burge S, Cabau C, Carlson JW, Chelala C, Chrysostomou C, Cittaro D, Collin O, Cordova R, Cutts RJ, Dassi E, Di Genova A, Djari A, Esposito A, Estrella H, Eyras E, Fernandez-Banet J, Forbes S, Free RC, Fujisawa T, Gadaleta E, Garcia-Manteiga JM, Goodstein D, Gray K, Guerra-Assuncao JA, Haggarty B, Han DJ, Han BW, Harris T, Harshbarger J, Hastings RK, Hayes RD, Hoede C, Hu S, Hu ZL, Hutchins L, Kan Z, Kawaji H, Keliet A, Kerhornou A, Kim S, Kinsella R, Klopp C, Kong L, Lawson D, Lazarevic D, Lee JH, Letellier T, Li CY, Lio P, Liu CJ, Luo J, Maass A, Mariette J, Maurel T, Merella S, Mohamed AM, Moreews F, Nabihoudine I, Ndegwa N, Noirot C, Perez-Llamas C, Primig M, Quattrone A, Quesneville H, Rambaldi D, Reecy J, Riba M, Rosanoff S, Saddiq AA, Salas E, Sallou O, Shepherd R, Simon R, Sperling L, Spooner W, Staines DM, Steinbach D, Stone K, Stupka E, Teague JW, Dayem Ullah AZ, Wang J, Ware D et al (2015) The BioMart community portal: an innovative alternative to large, centralized data repositories. *Nucleic Acids Res* 43:W589–W598. doi:[10.1093/nar/gkv350](https://doi.org/10.1093/nar/gkv350)
- Somel M, Khaitovich P, Bahn S, Paabo S, Lachmann M (2006) Gene expression becomes heterogeneous with age. *Curr Biol* 16:R359–R360. doi:[10.1016/j.cub.2006.04.024](https://doi.org/10.1016/j.cub.2006.04.024)
- Stegle O, Parts L, Durbin R, Winn J (2010) A Bayesian framework to account for complex non-genetic factors in gene expression levels greatly increases power in eQTL studies. *PLoS Comput Biol* 6:e1000770. doi:[10.1371/journal.pcbi.1000770](https://doi.org/10.1371/journal.pcbi.1000770)
- Stevens HE, Smith KM, Maragnoli ME, Fagel D, Borok E, Shanabrough M, Horvath TL, Vaccarino FM (2010) Fgfr2 is required for the development of the medial prefrontal cortex and its connections with limbic circuits. *J Neurosci* 30:5590–5602. doi:[10.1523/JNEUROSCI.5837-09.2010](https://doi.org/10.1523/JNEUROSCI.5837-09.2010)
- Tabuchi K, Blundell J, Etherton MR, Hammer RE, Liu X, Powell CM, Sudhof TC (2007) A neuroligin-3 mutation implicated in autism increases inhibitory synaptic transmission in mice. *Science* 318:71–76. doi:[10.1126/science.1146221](https://doi.org/10.1126/science.1146221)
- Tammimies K, Marshall CR, Walker S, Kaur G, Thiruvahindrapuram B, Lionel AC, Yuen RK, Uddin M, Roberts W, Weksberg R, Woodbury-Smith M, Zwaigenbaum L, Anagnostou E, Wang Z, Wei J, Howe JL, Gazzellone MJ, Lau L, Sung WW, Whitten K, Vardy C, Crosbie V, Tsang B, D'Abate L, Tong WW, Luscombe S, Doyle T, Carter MT, Szatmari P, Stuckless S, Merico D, Stavropoulos DJ, Scherer SW, Fernandez BA (2015) Molecular Diagnostic Yield of Chromosomal Microarray Analysis and Whole-Exome Sequencing in Children With Autism Spectrum Disorder. *JAMA* 314:895–903. doi:[10.1001/jama.2015.10078](https://doi.org/10.1001/jama.2015.10078)
- Tierney E, Bukelis I, Thompson RE, Ahmed K, Aneja A, Kratz L, Kelley RI (2006) Abnormalities of cholesterol metabolism in autism spectrum disorders. *Am J Med Genet B Neuropsychiatr Genet* 141B:666–668. doi:[10.1002/ajmg.b.30368](https://doi.org/10.1002/ajmg.b.30368)
- Turner TN, Hormozdiari F, Duyzend MH, McClymont SA, Hook PW, Iossifov I, Raja A, Baker C, Hoekzema K, Stessman HA, Zody MC, Nelson BJ, Huddleston J, Sandstrom R, Smith JD, Hanna D, Swanson JM, Faustman EM, Bamshad MJ, Stamatoyannopoulos J, Nickerson DA, McCallion AS, Darnell R, Eichler EE (2016) Genome Sequencing of Autism-Affected Families Reveals Disruption of Putative Noncoding Regulatory DNA. *Am J Hum Genet* 98:58–74. doi:[10.1016/j.ajhg.2015.11.023](https://doi.org/10.1016/j.ajhg.2015.11.023)
- Verboven S, Hubert M (2005) LIBRA: a MATLAB library for robust analysis. *Chemometr Intell Lab Syst* 75:127–136. doi:[10.1016/j.chemolab.2004.06.003](https://doi.org/10.1016/j.chemolab.2004.06.003)
- Voineagu I, Wang X, Johnston P, Lowe JK, Tian Y, Horvath S, Mill J, Cantor RM, Blencowe BJ, Geschwind DH (2011) Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature* 474:380–384. doi:[10.1038/nature10110](https://doi.org/10.1038/nature10110)
- Walsh P, Elsabbagh M, Bolton P, Singh I (2011) In search of biomarkers for autism: scientific, social and ethical challenges. *Nat Rev Neurosci* 12:603–612. doi:[10.1038/nrn3113](https://doi.org/10.1038/nrn3113)
- Wang K, Zhang H, Ma D, Bucan M, Glessner JT, Abrahams BS, Salyakina D, Imielinski M, Bradfield JP, Sleiman PM, Kim CE, Hou C, Frackelton E, Chiavacci R, Takahashi N, Sakurai T, Rappaport E, Lajonchere CM, Munson J, Estes A, Korvatska O, Piven J, Sonnenblick LI, Alvarez Retuerto AI, Herman EI, Dong H, Hutman T, Sigman M, Ozonoff S, Klin A, Owsley T, Sweeney JA, Brune CW, Cantor RM, Bernier R, Gilbert JR, Cuccaro ML, McMahon WM, Miller J, State MW, Wassink TH, Coon H, Levy SE, Schultz RT, Nurnberger JI, Haines JL, Sutcliffe JS, Cook EH, Minshew NJ, Buxbaum JD, Dawson G, Grant SF, Geschwind DH, Pericak-Vance MA, Schellenberg GD, Hakonarson H (2009) Common genetic variants on 5p14.1 associate with autism spectrum disorders. *Nature* 459:528–533. doi:[10.1038/nature07999](https://doi.org/10.1038/nature07999)
- Wang G, Yang E, Brinkmeyer-Langford CL, Cai JJ (2014) Additive, epistatic, and environmental effects through the lens of expression variability QTL in a twin cohort. *Genetics* 196:413–425. doi:[10.1534/genetics.113.157503](https://doi.org/10.1534/genetics.113.157503)
- Waterhouse L, Gillberg C (2014) Why autism must be taken apart. *J Autism Dev Disord* 44:1788–1792. doi:[10.1007/s10803-013-2030-5](https://doi.org/10.1007/s10803-013-2030-5)
- Weiss LA, Arking DE, Gene Discovery Project of Johns Hopkins and the Autism Consortium, Daly MJ, Chakravarti A (2009) A genome-wide linkage and association scan reveals novel loci for autism. *Nature* 461:802–808. doi:[10.1038/nature08490](https://doi.org/10.1038/nature08490)
- Willsey AJ, State MW (2015) Autism spectrum disorders: from genes to neurobiology. *Curr Opin Neurobiol* 30:92–99. doi:[10.1016/j.conb.2014.10.015](https://doi.org/10.1016/j.conb.2014.10.015)
- Willsey AJ, Sanders SJ, Li M, Dong S, Tebbenkamp AT, Muhle RA, Reilly SK, Lin L, Fertuzinhos S, Miller JA, Murtha MT, Bichsel C, Niu W, Cotney J, Ercan-Sencicek AG, Gockley J, Gupta AR, Han W, He X, Hoffman EJ, Klei L, Lei J, Liu W, Liu L, Lu C, Xu X, Zhu Y, Mane SM, Lein ES, Wei L, Noonan JP, Roeder K, Devlin B, Sestan N, State MW (2013) Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. *Cell* 155:997–1007. doi:[10.1016/j.cell.2013.10.020](https://doi.org/10.1016/j.cell.2013.10.020)
- Zeng Y, Wang G, Yang E, Ji G, Brinkmeyer-Langford CL, Cai JJ (2015) Aberrant gene expression in humans. *PLoS Genet* 11:e1004942. doi:[10.1371/journal.pgen.1004942](https://doi.org/10.1371/journal.pgen.1004942)