Title

Blood transcriptomic comparison of individuals with and without autism spectrum disorder: A combined-samples mega-analysis

Journal

American Journal of Medical Genetics Part B Neuropsychiatric Genetics, 174(3)

ISSN

1552-4841

Authors

Tylee, Daniel S

Hess, Jonathan L

Quinn, Thomas P

et al.

Publication Date

2017-04-01

DOI

10.1002/ajmg.b.32511

Published in final edited form as:

Am J Med Genet B Neuropsychiatr Genet. 2017 April ; 174(3): 181–201. doi:10.1002/ajmg.b.32511.

Blood Transcriptomic Comparison of Individuals with and without Autism Spectrum Disorder: A Combined-Samples Mega-Analysis

Daniel S. Tylee, B.A.[1], **Jonathan L. Hess, B.S.**[1], **Thomas P. Quinn, B.S.**[1], **Rahul Barve, M.D.**[1], **Hailiang Huang, Ph.D.**[2,3], **Yanli Zhang-James, M.D., Ph.D.**[1], **Jeffrey Chang, M.D.**[4], **Boryana S. Stamova, Ph.D.**[5], **Frank R. Sharp, M.D.**[5], **Irva Hertz-Picciotto, M.P.H., Ph.D.**[6], **Stephen V. Faraone, Ph.D.**[1,7], **Sek Won Kong, M.D.**[8], and **Stephen J. Glatt, Ph.D.**[1,†]

[1]Psychiatric Genetic Epidemiology & Neurobiology Laboratory (PsychGENe Lab); Departments of Psychiatry and Behavioral Sciences & Neuroscience and Physiology; SUNY Upstate Medical University; Syracuse, NY, U.S.A

[2]Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts, USA

[3]Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA

[4]Department of Psychiatry and Behavioral Sciences, SUNY Downstate Medical Center, Brooklyn, NY, U.S.A

[5]Department of Neurology, UC Davis School of Medicine, Sacramento, CA, USA

[6]Department of Public Health Sciences and UC Davis MIND Institute, School of Medicine, Davis, CA

[7]K.G. Jebsen Centre for Research on Neuropsychiatric Disorders, University of Bergen, Bergen, Norway

[8]Computational Health Informatics Program, Boston Children's Hospital; Department of Pediatrics, Harvard Medical School, Boston, MA, U.S.A

Abstract

Blood-based microarray studies comparing individuals affected with autism spectrum disorder (ASD) and typically developing individuals help characterize differences in circulating immune cell functions and offer potential biomarker signal. We sought to combine the subject-level data from previously published studies by mega-analysis to increase the statistical power. We identified studies that compared *ex-vivo* blood or lymphocytes from ASD-affected individuals and unrelated comparison subjects using Affymetrix or Illumina array platforms. Raw microarray data and clinical meta-data were obtained from seven studies, totaling 626 affected and 447 comparison subjects. Microarray data were processed using uniform methods. Covariate-controlled mixed-effect linear models were used to identify gene transcripts and co-expression network modules that

[†]To whom correspondence should be addressed: SUNY Upstate Medical University, 750 East Adams Street, Syracuse, NY 13210, Phone: (315) 464-7742, stephen.glatt@psychgenelab.com.

Supplementary Materials are available at The American Journal of Medical Genetics Part B: Neuropsychiatric Genetics website.

were significantly associated with diagnostic status. Permutation-based gene-set analysis was used to identify functionally related sets of genes that were over- and under-expressed among ASD samples. Our results were consistent with diminished interferon-, EGF-, PDGF-, PI3K-AKT-mTOR-, and RAS-MAPK-signaling cascades, and increased ribosomal translation and NK-cell related activity in ASD. We explored evidence for sex-differences in the ASD-related transcriptomic signature. We also demonstrated that machine-learning classifiers using blood transcriptome data perform with moderate accuracy when data are combined across studies. Comparing our results with those from blood-based studies of protein biomarkers (*e.g*., cytokines and trophic factors), we propose that ASD may feature decoupling between certain circulating signaling proteins (higher in ASD samples) and the transcriptional cascades which they typically elicit within circulating immune cells (lower in ASD samples). These findings provide insight into ASD-related transcriptional differences in circulating immune cells.

Additional Keywords

gene expression; microarray; immune system; machine learning

INTRODUCTION

The molecular bases of autism spectrum disorder (ASD) remain largely unresolved despite decades of research. This situation impedes progress toward biologically based risk assessment and diagnostic testing, early detection, and the identification or development of rationally selected therapeutics aimed at improving developmental trajectories and functioning. As such, the molecular correlates of ASD have been pursued at multiple levels. Both twin- and common variant SNP-based heritability studies suggest that a substantial proportion of risk for developing ASD appears to be mediated by genetic factors (Sandin et al., 2014; Gaugler et al., 2014). The largest meta-analytic genome wide association (GWA) study to date identified one signal that surpassed a genome-wide significance threshold for discovery (www.med.unc.edu/pgc/downloads); yet many others appear poised to rise to this level of significance with larger sample sizes. However, any given risk-associated common variants likely impart a relatively small (< 1%) increase in absolute risk. Alternatively, studies of rare variants (single nucleotide, structural and copy-number variants, or even larger chromosomal abnormalities) indicate that these variants may impart a considerably larger increase in an individual's absolute risk for ASD (Levy et al., 2011; Sanders et al., 2011, 2012). Some have suggested that hundreds of distinct rare variants may contribute to ASD risk in the population (Levy et al., 2011), but that any specific variant may be found in < 1 % of cases. Taken together, these data suggest that ASD arises from heterogeneous underlying architectures of genetic and environmental risk factors, with different genetic variants (common and rare) contributing to disorder in different individuals or combining in complex ways within and across individuals and via interactions with the broad range of non-genetic (*i.e*., environmental) factors, sometimes referred to as the 'exposome'. In light of these observations, much effort has been made to identify whether the apparently distinct genetic risk factors tend to converge into one or more unifying pathophysiological mechanisms (De Rubeis et al., 2014; Geschwind, 2008), which could also assist in identification of candidate exposures (Stamova et al., 2011; Tian et al., 2011) that influence those same molecular pathways. Examining aspects of biology lying downstream from both genetic and environmental factors – *e.g.*, genome-wide gene expression levels (transcriptome) – may help provide a useful framework for understanding of the developmental pathophysiology of ASD.

In the past 15 years, the ASD transcriptome has received considerable attention with the emergence of microarrays that enable unbiased and simultaneous surveys of most human genes. Studies of *post-mortem* brain tissues are best suited to help us understand pathophysiology, but these have been relatively rare, often relied on a small number of samples, and they have examined various brain regions. Garbett *et al.* (2008) reported that transcriptomic differences in ASD temporal cortex were enriched with inflammatory activity (*e.g.* IL-1/2/6, toll-like receptors (TLRs), NF-κB), cell cycle and EGF receptor signaling cascades, among others; this analytical approach did not allow for directional conclusions to be made about the enriched ontologies. Voineagu *et al.* (2011) reported under-expression of frontal/temporal cortex transcriptional network modules enriched with genes involved in synaptic function, vesicular transport, and neuronal projection, as well as ASD candidate-genes. They also reported over-expression of a module enriched with astrocyte and activated microglial markers, and genes involved in immune and inflammatory responses, among others; these authors suggested that inflammatory responses may be secondary to synaptic abnormalities. Chow *et al.* (2012) profiled prefrontal cortical tissues. Among samples from younger subjects (≤ 14 years), they identified dysregulation of genes enriched in DNA damage responses, cell cycle, neurogenesis and neurodevelopment, apoptosis, and inflammatory responses. Among samples from older subjects (>14 years), they observed enrichment for genes involved in cellular differentiation, development, mitogenic signaling, apoptosis, oxidative stress, and tissue repair and remodeling. Whether these emergent functions were predominantly over-expressed or under-expressed in ASD samples was difficult to discern based on the analytical approach, with apparent evidence for mixed effects among the gene transcripts. Ginsberg *et al.* (2012) profiled occipital cortex and reported under-expression of mitochondrial and protein synthesis genes. They also examined a subgroup of cerebellar samples, which showed differential expression of genes involved in NF-κB related inflammatory signal transduction, when compared with unaffected subjects; network-level analyses of cerebellar tissues also implicated groups of genes over-representing functions including inflammatory and purinergic signaling, as well as myelination; these functional enrichment analyses did not support clear directional conclusions. Gupta *et al.* (2014) performed the largest published RNA sequencing study on samples of frontal, pre-frontal and occipital cortices; functional enrichment of dysregulated genes and ASD-associated network modules implicated over-expression of genes enriched with markers of the M2 microglial phenotype, cellular antiviral responses, and type I and II interferon signaling, as well as other down-regulated functions, including neurotransmission, GABAergic signaling and hormone signaling. Because studies of *post-mortem* brain samples typically reflect a range of ages and generally do not capture the early developmental time windows during which ASD-related differences in brain development are first occurring, it remains unclear whether these signals reflect causal aspects of ASD pathophysiology or instead reflect correlational signals secondary to pathophysiology or treatment. Collectively, these studies appear to support the idea that transcripts involved in immunologic functions are over-expressed in the ASD brain, while transcripts involved in neurodevelopment and synaptic signaling are under-expressed.

Studies of blood have also been pursued in efforts to identify biomarkers using readily assessable tissue and to shed light on differences in the ASD immunologic milieu. Gregg *et al.* (2008) identified natural killer (NK)-cell cytotoxicity as an emergent function among dysregulated genes in whole blood samples from children with ASD. Enstrom *et al.* (2009) focused their analyses on ASD cases with high NK-cell gene expression, comparing them against typically developing children with low NK-cell gene expression, and found MHC class I-mediated antigen presentation as an enriched function among up-regulated genes, whereas down-regulated genes were enriched with ribosomal protein synthetic and cell metabolic functions. Alter *et al.* (2011) examined transcriptomic differences in isolated lymphocytes, reporting that the preponderance of dysregulated genes were down-regulated in ASD cases (and also in typically developing children with older fathers), as compared with younger-fathered comparison subjects, and were enriched for zinc-binding, transcription factor, and ubiquitin ligase activity. Kuwano *et al.* (2011) examined whole blood samples from young adults with ASD; qualitative functional analysis of the top hits implicated cell morphology, cellular assembly and organization, nervous system development and function. In a relatively large sample of children, Kong *et al.* (2012) identified dysregulation of genes subserving trophic (*e.g.* neurotrophins, EGF/ErbB and VEGF), inflammatory, and cytoskeletal signaling pathways; notably many of the pathways shared common members of the MAP kinase gene family. Glatt *et al.* (2012) identified functional enrichment of immunologic, hemoglobin complex, and nucleotide-binding genes among those selected by a machine-learning diagnostic classifier. A second study by Kong *et al.* (2013) primarily compared probands and unaffected siblings; they identified up-regulation of ribosomal and spliceosomal gene-sets, and down-regulation of genes involved in neuroactive ligand receptor signaling, calcium signaling, and gap junctions. Differences in profiling, analytical approaches and reporting of results makes it challenging to infer whether emergent biological themes are consistently dysregulated across these studies, thus underscoring the need for systematic re-analysis.

Blood-based investigations could afford access to biomarker signals at critical postnatal developmental time points, (Glatt et al., 2012; Kong et al., 2012) yet these studies have typically (1) included subjects from a range of ages that extend beyond key developmental windows, (2) been underpowered to confidently detect dysregulated genes, and (3) have not consistently assessed directionality when examining dysregulated emergent functions. As such, two recent studies have sought to consolidate the knowledge of transcriptomic abnormalities in ASD *via* meta-analysis. Ning *et al.* (2015) applied uniform preprocessing methods to data reflecting blood and brain microarray studies and combined summary statistics across samples to identify dysregulated genes with improved confidence. Their enrichment analyses implicated ribosomal translation and gene transcription (and their regulation), as well as immunologic functions (MHC class I, T-cell selection and activation, cytokine signaling), fatty acid metabolism, and anti-apoptotic genes. While this study afforded improved statistical power, it did not model the effects of potentially influential covariates of gene expression, it provided relatively little insight into the directionality of changes in functional gene-sets, and did not specify whether effects were shared between tissues or were tissue-specific. Ch'ng *et al.* (2015) also performed a well-powered meta-analysis of array-based studies of blood, transformed lymphoblastoid cells, and brain. However, their conservative approach focused only on genes that showed a consistently significant effect across all constituent studies for a given tissue; as such, they did not identify significant functional enrichment among transcripts consistently dysregulated in the blood.

In comparison to meta-analysis, mega-analysis offers an alternative strategy for deriving consensus across multiple studies; this involves pooling of individual-level clinical and normalized biological data from multiple studies and modeling with appropriate correction for between-study variations (Seifuddin et al., 2013; Mistry et al., 2013), as well as statistical controls for factors that are consistently reported across studies (*i.e.*, sex, age), the potential to control for latent covariance structures (*i.e.* surrogate variables) that may differ across studies or diagnostic groups (Leek and Storey, 2007; Stegle et al., 2012), and the ability to tolerate missing data (*e.g.*, clinical covariates and genes reported in only a subset of studies). This approach also easily lends itself to gene co-expression network analyses (described below). We sought to mega-analyze microarray data generated on popular platforms (Illumina and Affymetrix), reflecting *ex-vivo* whole blood or non-transformed leukocytes from ASD-affected individuals and unrelated comparison subjects. We hypothesized that this approach would identify ASD-associated transcripts surpassing rigorous statistical correction and that co-expression network and ontology-based analyses would identify over-expression of groups of genes subserving immunologic and trophic signaling functions.

METHODS

Literature Search and Study Selection

The annotated MOOSE guidelines for this mega-analysis of observational studies are provided in Supplementary Table 1. Selected studies and respective clinical and demographic features are shown in Table 1; excluded studies (along with the rationale for exclusion) are shown in Supplementary Table 2. We performed a literature search (SCOPUS) and microarray database searches (NCBI GEO and EMBL-EBI ArrayExpress) for microarray-based studies of whole blood- or leukocyte-based gene expression in subjects with autism or ASD, and including a control group composed of typically developing, unrelated comparison subjects. For full search terms, see Supplementary Table 1. We only included studies that were generated on popular microarray platforms (Illumina and Affymetrix) in order to allow consistent pre-processing with publicly available software and reliable mapping of probes to HGNC symbols. We sought to characterize transcriptomic differences associated with ASD using only *ex-vivo* peripheral blood samples or isolated leukocyte samples derived from peripheral blood. We chose to exclude studies that examined transformed (*i.e.* lymphoblastoid) cell lines because it is unclear the extent to which these cells reflect features of the circulating immunologic milieu in living volunteers. We also chose to exclude studies that compared ASD cases with unaffected members of the same family, because of the possibility that heritable genetic factors, environmental risk factors and/or intermediate phenotypes (*e.g.*, gene expression profiles, neurocognitive features, personality features, *etc.*) may be shared among family members discordant for ASD diagnostic status, which might mask the detection of transcriptomic differences that would otherwise be observed when examining unrelated comparison subjects. Seven studies meeting these preliminary criteria were identified (Table 1). We obtained microarray and clinical covariate data from the corresponding authors of the original studies or from publicly available data repositories, ensuring to the best of our ability that no subjects were included in multiple studies. Because we wanted to maximize the sample size, we accepted the diagnostic criteria used by each individual study site, some of which were based on clinician assessments and others based on standardized screening tools. We did not apply any additional filtering of subjects based on medical comorbidities beyond what was described by the authors of the previous studies who provided the samples (see Table 1 for brief descriptions and full citations of the original publications).

Coding and Harmonization of Clinical and Demographic Information

We obtained clinical and demographic covariate data from public data repositories and, when necessary, from corresponding authors of the previously published articles. In order to facilitate explicit modeling of these effects, we sought to harmonize the clinical demographic information across different study sites through re-coding. A variable for Study ID was created and separate levels were created for instances where a single publication included data generated from two different array platforms (as displayed in Table 1). Diagnostic labels (as determined by the authors of the original studies, including autism, autism spectrum disorder, Asperger's disorder, and pervasive developmental disorder not otherwise specified) were recoded as “ASD” and unaffected subjects were uniformly recoded as “Comparison”. This decision is in keeping with the latest revision of the nosology within the DSM-V (American Psychiatric Association, 2013). Age was recoded in number of years for all subjects. Sample type was coded as either “whole blood” or “lymphocytes”. Data related to self-reported ancestry, ethnicity, and nationality required multistep harmonization because different studies provided this information in different formats. Most studies included only one column of information. For single-column ancestry data, responses including “Caucasian” or the mention of a specific European nationality (*e.g.* “Irish”) were uniformly recoded as “European.” Using a similar approach, we examined whether broad categories of “Asian”, “Latin American” and “African” could be populated, but the resulting groups were observed to be relatively small in comparison to the “European” sample. The decision was made to preserve statistical power by combining these groups into a single category of “Non-European” ancestry. Subjects reporting a mix of both European and Non-European ancestry were recoded as missing data. Some studies reported two separate columns of ancestry-related information (*e.g.,* reporting “Race” [Caucasian, Black, Asian] separately from “Hispanic” [yes vs no]. For this scenario, we encoded each column of data separately using previously described rules. We then condensed the two columns into a single column, such that subjects reporting a combination of European and Non-European ancestry were coded as missing data.

Microarray Data Processing and Quality Control

All data processing and analyses were performed within the *R* statistical computing environment. Data from each study were processed and normalized independently, and received uniform quality control treatment. Affymetrix arrays underwent robust multi-array average (RMA) normalization(Irizarry et al., 2003), with additional GC-correction whenever possible (*e.g.*, not compatible with Affymetrix Human Exon 1.0ST array) using the *affy* (Gautier et al., 2004)*, oligo* (Carvalho and Irizarry, 2010), and *gcrma* (Wu and Irizarry, 2016) packages. Illumina data were imported as background-corrected expression values using default *Genome Studio* (Illumina Inc.; San Diego, California) methods. All chip sets were quantile-normalized and *log*2 transformed. Principal components analysis was performed for each study separately; samples were plotted along the first three components, and outlier subjects were removed if they deviated beyond the 4 *s.d.* ellipsoid extended from the grand mean. For the studies that provided batch information, the microarray data were visually inspected and batch effects were removed using the *ComBat* function of the *SVA* package (Leek et al., 2012). We mapped microarray probes to HGNC gene symbols using *biomaRt* (Durinck et al., 2009) and *AnnotationDbi* (Pages et al., 2016) and collapsed expression values by taking the median when multiple probes mapped to the same HGNC symbol. Finally, for each gene within each individual study, expression values were *z*-transformed in order to normalize the range and variance of expression across datasets generated on different array platforms. *Z*-transformed data from each study were then combined (based on common HGNC symbols) across studies to create one large matrix of expression and covariate data for mega-analysis (n = 1,073; as shown in Table 1).

Leukocyte Stratification Analysis

Complete blood cell counts with leukocyte differentials were not available for the subjects examined in our study. In order to explore potential ASD-related differences in the proportions of leukocyte subtypes between ASD cases and unaffected comparison subjects (which could influence the transcriptomic differences between diagnostic groups, due to the different transcriptomes of each cell type), we performed microarray de-convolution analysis using previously described methods (Gaujoux and Seoighe, 2013); this approach examines the expression of genes that are highly specific for each class of leukocyte and computes surrogate values for each subject. Diagnostic group differences in the surrogate values were compared using an independent samples *t*-tests with family-wise Bonferroni correction for multiple testing. However, because ASD-related differences in the proportion of circulating leukocyte subtypes could be an important driver of differences in the peripheral immune milieu, and because statistically controlling for such effects could undermine our ability to characterize ASD-related transcriptomic differences, we did not include the variables generated from de-convolution analysis as covariates in the linear models used to detect differential expression (described below).

Mixed-Effect Linear Models for Single-Gene Analyses

Mega-analysis was performed on the combined data matrix using mixed-effect linear modeling, as implemented within the *lmerTest* package (Kuznetsova et al., 2016). This approach subsets the data matrix for non-missing expression and covariate values and employs the Satterthwaite method to approximate the effective degrees of freedom for a linear combination of independent sample variances for each transcript. We examined the effect of diagnostic status (ASD, comparison), and included covariates for age (continuous), self-reported ancestry (European, Non-European), sex (female, male), and sample-type (whole blood, lymphocytes). A random effect of study ID was included to control for study-specific effects on expression. While modeling the expression of each transcript, step-wise removal of non-significant covariates ($p > 0.10$) was performed to maximize statistical power. A total of 21,968 gene-level mixed-effect models converged to produce valid statistical results in the full sample ($n = 1{,}073$). For multiple-test correction, we examined Bonferroni-corrected *p*-values in order to conservatively define differentially expressed genes. Additionally, we also utilized a more permissive False Discovery Rate (FDR) *q*-value (Storey, 2003) to control the family-wise error rate at 5%, in order to allow more transcripts to move forward for GWAS and candidate-gene enrichment analyses. For each gene, the effect sizes derived from these covariate-controlled mixed-effect models were supplied for permutation-based gene-set analysis (described below).

Secondarily, we also performed separate analyses for each sex, in order to explore sex-differences in the blood transcriptomic signature of ASD. For the sex-specific single-gene analyses, similar mixed-effect models (without sex specified as a covariate) were used to identify dysregulated genes separately for females and males. Among females ($n_{\mathrm{ASD}} = 90$, $n_{\mathrm{comparison}} = 117$), 19,508 gene-level mixed-effect models converged to produce valid statistical results. Among males ($n_{\mathrm{ASD}} = 536$, $n_{\mathrm{comparison}} = 330$), 17,268 gene-level mixed-effect models converged to produce valid results. All cross-sex comparisons of dysregulated genes (*i.e.*, hypergeometric tests of gene-list overlap) were predicated on the common background of genes analyzed (genes = 17,268). For each gene, the effect sizes derived from these covariate-controlled mixed-effect models were supplied for permutation-based gene-set analysis (described below).

Single-Gene Meta-Analysis and Sensitivity Analysis

For genes reaching a Bonferroni-corrected level of significance in the mega-analysis of the full sample, we performed traditional meta-analysis to explore heterogeneity across studies, as well as sensitivity analysis, to explore whether certain studies contributed disproportionately to the findings. The meta-analyses used linear modeling on each individual sample (as shown in Table 1), without mixed-effects and without step-wise removal of covariates. For each gene transcript, we performed inverse variance weighted meta-analysis using the *meta* package (Schwarzer, 2016) to combine single-study results and compare these with mega-analysis summary statistics. We created forest plots to depict this information. For heterogeneity analysis, we withheld each individual study and repeated the mixed-effect linear modeling. We describe these results briefly in the main text, with accompanying Supplementary Tables and Supplementary Figures.

Gene-Set Analysis based on Permutation of Single-Gene Statistics

Permutation-based gene-set analysis was performed using the *Piano* package (Väremo et al., 2013); for this analysis we supplied the summary statistics from the covariate-controlled mega-analysis of the full sample (*i.e.*, covariate-adjusted diagnostic group differences) to assess whether *a-priori*-defined gene-sets reflecting various functional and ontological themes show evidence of differential expression compared to randomly resampled gene-sets of equal size. This approach to functional enrichment analysis does not require an arbitrary gene-list threshold (typically based on *p-value*) and can identify over- and under-expressed functional sets, even if the individual genes subserving those functions fail to reach a threshold of differential expression. Specifically, for each gene within each *a-priori-defined* gene-set, we assessed the covariate-adjusted diagnostic group mean difference and these values were combined using the arithmetic mean to derive an average ASD-related difference for the gene-set. This value was then compared against 1×10^6 randomly selected gene-sets of equal size in order to generate empirical *p*-values for tests of five distinct hypotheses. For the non-directional test of dysregulation, the mean *p*-value of the target set was compared to the mean *p*-value of the randomly permuted sets. For tests of an absolute directional hypotheses (*i.e.* all up-regulated and all down-regulated), the mean group difference of the target set was compared against the mean for permuted sets. For the test of a mixed directional hypothesis, each target gene-set was subset to include only genes with positive (*i.e.* for mixed up-regulated effects) or negative (i.e. for mixed down-regulated effects) diagnostic group mean differences; these values were compared against permuted mean differences drawn from the background reflecting all same-signed test statistics. Seven gene-set databases were obtained from the Molecular Signature Database (Broad Institute): H: hallmark gene-sets, C1: positional gene-sets, C2: curated gene-sets, C3: motif gene-sets, C5: GO gene-sets, and C6: oncologic signatures, and C7 immunologic signatures; together these sets capture collective knowledge of gene participation in functional pathways, known and predicted regulatory relationships, and chromosomal locations. The names of these gene sets (as shown in Supplementary Tables) can be queried within the Database's website (http://software.broadinstitute.org/gsea/msigdb) in order to learn more about exactly how the gene-set was derived. Only gene-sets intersecting with the target list of test statistics (with 4 to 500 genes) were analyzed. Multiple test correction was performed separately for each combination of database and test hypothesis using the Bonferroni method to control the database-wise error rate at 5%. For gene-sets where multiple test hypotheses were significant (*i.e.* non-directional, mixed-up-regulated and all up-regulated effects), we elected to report the directional effects, such that all up-regulated was given the highest preference. We also combined the results of mixed directional and absolute directional tests to facilitate cross-sex comparison. We observed no instances of conflicting directional tests (*e.g.* both up- and down-regulated hypotheses reaching significance). This analysis was repeated for using the results of each sex-specific covariate-controlled mega-analysis and ASD-associated gene-sets were compared across sexes.

Gene Co-Expression Network Analyses

We used the weighted gene co-expression network analysis (*WGCNA*) package to assess networks of correlated genes (Langfelder et al., 2011). This analysis was performed on the subset of Z-transformed expression data reflecting genes that were reported across all studies (genes = 14,931). We utilized both network preservation-type and module eigengene-type approaches. Both types of WGCNA analyses are predicated on the assignment of individual gene transcripts to network modules (*i.e.*, groups of transcripts whose expression levels are correlated across the samples). Networks modules were detected based on Pearson correlation coefficients for all pairs of genes using an unsigned approach; absolute correlation coefficients between all gene pairs were raised to the power β, which was selected as the lowest soft-threshold power that approximated a scale-free topology network (β = 6). Additionally, we set the following parameters to construct the networks (kept default if not specified) within the *blockwiseModules* command: *deepSplit* = 3, *TOMType* = "unsigned", *minModuleSize* = 60, *minCoreKME* = 0.5, *minCoreKMESize* = 20, *minKMEtoStay* = 0, *reassignThreshold* = 1×10^{-6}, *mergeCutHeight* = 0.25, *detectCutHeight* = 0.995, and *maxBlockSize* = 5000.

For the preservation-type analysis, correlation structures are identified within a reference sample (*e.g.*, unaffected comparison subjects), and then these relationships are tested for preservation within another sample (*e.g.*, affected subjects), though this analysis can also be repeated in the opposite direction. The extent to which module correlation structure is preserved can be inferred from the *z*-summary statistic, with *z*-summary ≥ 10 generally considered highly preserved, 10 > z-summary ≥ 2 considered moderately preserved, and z-summary < 2 considered poorly preserved. When a module is poorly preserved in the tested sample, it implies that the correlations among genes are different than those captured in the reference sample. Preservation analyses were run bi-directionally (*i.e.*, two separate runs, specifying the comparison individuals, then the ASD-affected individuals, as the reference group).

We also performed a module eigengene analysis. This type of analysis assumes that the correlation structures are similar between the two groups, and it summarizes the expression levels for large sets of highly correlated genes as a module eigengene value (similar to a principal component score). These module eigengene values were calculated for each subject, and we sought to identify ASD-associated modules based on the covariate-controlled mixed-effect models with step-wise removal of non-significant covariates (described previously for the single-gene analyses). For a given module, if the eigengene values show a significant between-group difference, it implies that the expression of genes included within the module, when examined together as a network function, show evidence for differential expression. Linear model test statistics were corrected for the total number of network modules using the Benjamini-Hochberg method and modules robustly associated with diagnostic status were further characterized. For ASD-associated modules, we identified highly connected hub genes and assessed functional and cell-type enrichment (described below). In order to explore sex-differences in ASD-related co-expression networks, we repeated both preservation and module eigengene analyses separately for females and males, comparing the findings across sexes.

Enrichment Analysis for Biological Annotations, Cell-Type Markers, and GWAS Signal

For gene lists corresponding to ASD-associated network modules, functional enrichment analysis was performed using hypergeometric tests to assess the degree of overlap between each gene list and the contents of the Molecular Signature Database (H, C1, C2, C3, C5, C6 and C7) collections. When testing a given gene list, any gene-set that contained at least one overlapping gene was retained, and then family-wise BH *p*-value correction was applied for the total number of surviving gene-sets per collection. However, we only reported significantly enriched gene-sets (BH $p < 0.05$) sharing at least two genes with the target gene list. The same rules were applied when hypergeometric testing for enrichment of cell-specific markers; these lists were created by combining those supplied within the *CellMix* package (Gaujoux and Seoighe, 2013) with those identified in an independent study

(Watkins et al., 2009). In order to assess whether gene-lists or network modules were enriched with genome-wide association signal, we obtained the gene-level results from the recent Psychiatric Genomics Consortium (PGC) autism meta-analysis (Hailiang Huang, personal communication); these were generated for 18,533 genes using the FAST algorithm (Chanda et al., 2013), which uses a greedy search algorithm to aggregate linkage disequilibrium-independent signals and applies correction for gene length (Huang et al., 2011). For qualitative reporting of PGC Autism GWAS hits within the gene lists resulting from our transcriptomic study, we highlight any gene that showed an uncorrected gene-level association of *p*-value < 0.05. Additionally, for each gene list of interest in the present study, we performed permutation-based testing to assess whether the lists showed quantitative evidence of enrichment with GWAS signal. For a given gene list, the aggregate PGC association signal was obtained by combining *p*-values across genes using Fisher's method. We then randomly resampled equal-sized lists of genes 10,000 times and summarized association *p*-values for these genes. Empirical *p*-values were calculated for the target gene lists to assess whether GWAS signal was significantly enriched. Finally, gene lists of interest were also compared against the unfiltered list of 706 ASD candidate-genes curated from multiple sources by SFARIgene.org (obtained 12/01/2014). For the purpose of depiction in Figures and Supplementary Figures, genes meeting both criteria as a GWAS hit and a SFARIgene candidate were preferentially reported as GWAS hits; full information is reported in Supplementary Tables.

Additionally, we examined the lists of dysregulated genes detected in the full sample (thresholded at both FDR-*q* < 0.05 and Bonferroni *p* < 0.05), as well as the genes participating in the ASD-associated (Benjamini-Hochberg p < 0.05) network modules using Ingenuity Pathway Analysis (IPA) Software (Redwood City, California, USA). Each of these gene lists was overlaid with summary statistics (covariate-adjusted mean differences between diagnostic groups) and were submitted for IPA's Core Analysis. We examined the Canonical Pathways enriched for each gene list and we identified pathways that showed consistent enrichment across all three gene lists. For Canonical Pathways of interest, we isolated the genes subserving these pathways. We then used the Molecule Activity Predictor (MAP) tool to predict and visualize the upstream and/or downstream effects based on the expression changes of genes in the pathway.

Machine Learning Classification using Blood Transcriptomic Data

We sought to construct and independently validate machine-learning classifiers using a custom-built *R* package (*exprso*; https://github.com/tpq/exprso), which integrates a number of existing packages, including *e1071* (Meyer et al., 2015), *limma* (Ritchie et al., 2015)*, Matching* (Sekhon, 2015)*, mRMRe* (De Jay et al., 2013)*, nnet* (Ripley and Venables, 2015)*, ROCR* (Sing et al., 2005)*, pathClass* (Johannes et al., 2011)*, penalizedSVM* (Becker et al., 2009) *and sampling* (Tillé and Matei, 2015). Code for our implementation of this classification pipeline can be shared upon request to the corresponding author. As in the network co-expression analysis, we utilized the set Z-transformed expression values for genes that were reported across all studies (genes = 14,931), to avoid problems created by missing data. All samples (*n* = 1,073) were pseudo-randomly divided into a training sample (67% of *n*); this sample was selected to have an equal number of affected and unaffected subjects with, on average, similar age between groups (*t*-test $p > 0.30$). The counterpart validation sample contained the remaining 33% of subjects. All feature selection and model fitting was performed within the training sample. The best performing models were selected based on performance within the training sample and they were combined into ensembles, which were then evaluated in the withheld validation sample. This procedure was repeated 20 times, averaging the accuracy measurements across each resulting validation set in order to ensure a robust estimate of classifier accuracy.

In order to select models that are robust, the training data were analyzed in a manner similar to nested 2-layer cross-validation. The training sample was pseudo-randomly bootstrapped five times, ensuring that at least 1 ASD-affected and 1 typically developing subject was included in each of the resulting partitions. Within each bootstrap, 67% of were samples randomly assigned for classifier fitting (boot67) and 33% were withheld from classifier fitting for the evaluation of generalizability (boot33). This created different versions of the boot67 sample across the 5 bootstraps. Within each boot67 sample, affected and unaffected subjects were compared for gene expression using various filter methods (*t*-test, Kolmogorov–Smirnov test, mRMRe [De Jay et al., 2013]) and the selected features were supplied to the machine learning algorithm. We employed support vector machines with linear kernels, random forests, and neural networks. Each feature selection method was paired with each learning machine over a grid of possible machine settings. Each model was evaluated using 10-fold cross-validation within boot67 (the sample used for model fitting) and was also evaluated based on the area under the receiver operating curve within the boot33 sample (withheld from classifier fitting). We evaluated the accuracy of models by combining the boot67 and boot33 accuracy measures as a multiplied product; this was done to provide an extra layer of protection against the selection of models over-fit to the boot67 sample. For each type of learning machine, we selected the top 3 models from each bootstrap, regardless of the feature selection employed. These prediction models were combined to create an ensemble whose combined prediction probability was used to assign samples into a diagnostic group; this ensemble was then evaluated within the completely withheld validation sample. Because we observed that the maximum withheld sample classification accuracy varied across different pseudo-randomly selected training/validation sets, we repeated this procedure 20 times, each time selecting a new training/validation set, in order to ensure a robust point estimate of maximum generalizable classification. For the best performing ensembles, we reported the list of genes and the frequency at which they were selected across the 20 runs.

Surrogate Variable Analysis (SVA) for Single-Gene and Gene-Set Analyses

In transcriptomic studies, it is possible that inadequately controlled factors (*e.g.,* RNA quality, stratification based on genetic ancestry or leukocyte subpopulations, history of recent infectious illness, *etc.*) could differ between cases and controls within or between studies, thus contributing to false positive findings of differentially expressed genes and their emergent functions. In order to guard against this possibility and ensure the robustness of the findings reported in the covariate-controlled mixed-effect mega-analysis, we sought to re-analyze our data using a second strategy called Surrogate Variable Analysis (SVA; Leek et al., 2012). For this secondary approach, we examined the subset of our sample described above for network-level and classification analyses (n = 1,073, genes = 14,931) in order to avoid problems posed by missing data. In order to be very conservative, we examined the full sample and performed principal components analysis; we identified an additional 40 subjects that could be considered outliers based on their distance (> 4 *s.d.*) from the combined-sample mean for the first 5 principal components, leaving a total of 1,033 samples. We then used the *SVA* package (Leek et al., 2012) with Leek's method to identify the optimal number of surrogate variables ($k = 4$) while preserving the transcriptomic patterns related to diagnostic status. For each gene transcript, the effects of these 4 surrogate variables were removed from the expression values using linear model residuals. Single-gene mega-analysis was then repeated using mixed-effect linear modeling, however only the random effect (study ID) and a single fixed effect (diagnostic status) were included in the models and no step-wise removal was performed. For each gene, the resulting effect sizes were supplied for permutation-based gene-set analysis as described previously. Full sets of results files, summary statistics, and *R* scripts can be provided upon request to the corresponding author.

RESULTS

Leukocyte Stratification Analysis

Because whole blood samples were used in the majority of studies, we examined whether expression differences in leukocyte subtype-specific markers differed between affected individuals and comparison subjects. The results of microarray deconvolution analysis were consistent with increased expression of genes specific to NK cells and T-helper cells, and decreased expression of genes specific to activated dendritic cells in ASD samples (*t*-test Bonferroni $p < 0.03$). These data suggest that ASD might feature differences in the abundance of these leukocyte subtypes, which typically reflect only a minority of circulating leukocytes; differential leukocyte counts were not available for our samples, so these findings could not be confirmed.

ASD-Associated Transcripts

Next we examined differential expression of single genes in association with diagnostic status. From our covariate-controlled linear mixed models, 90 transcripts were dysregulated at a conservative Bonferroni-corrected $p < 0.05$ (Supplementary Table 3, bold gene symbols); among these, 66 were over-expressed and 24 were under-expressed in ASD (two-tailed sign test $p < 5.4\times10^{-6}$). A total of 1,572 reached a more liberal FDR *q*-value < 0.05, with 796 over-expressed and 776 under-expressed (Supplementary Table 3). Among these 1,572 transcripts, we did not observe significant enrichment with gene-level GWAS signal using a permutation-based analysis ($p = 0.092$); eighty-four of these genes showed at least nominal GWAS association (gene-level uncorrected $p < 0.05$) in a recent GWAS meta-analysis (Hailiang Huang, personal communication). Additionally, we identified forty as SFARIgene candidates; all GWAS hits and SFARI genes are indicated in Supplementary Table 3.

Mixed effect modeling was repeated to assess the effects of diagnostic status using data that was residualized for latent sources of variance. Using SVA-residualized expression values, 14,929 gene transcripts produced a valid statistical result (Supplementary Table 3). Among these, ninety-two reached a conservative Bonferroni $p < 0.05$ (corrected for 14,929 tests) and 1,564 reached a more liberal FDR q-value < 0.05 (Supplementary Table 3). We then sought to examine the ninety genes that reached a Bonferroni-corrected $p < 0.05$ in the primary covariate-controlled analysis of non-residualized data to determine whether the effects were robust within the SVA-corrected re-analysis. Among the ninety genes, 52 were consistently reported across studies, thus were available for analysis in the SVA-residualized data (Supplementary Table 3), and among these, 100% showed significant association with diagnostic status at FDR $q < 0.05$ (mean uncorrected p-value = 2.4×10^{-5}, maximum p-value = 5.3×10^{-4}), and 75% showed a significant association after Bonferroni-correction for 14,929 tests. All 52 genes showed a consistent direction of effect across both analyses. Similarly, among the 1,571 genes reaching an FDR $q < 0.05$ in the primary analysis, 1059 were analyzed in the SVA-residualized data, with 75% also reaching an FDR $q < 0.05$ and all of these showing a consistent direction of effect when compared with the primary analysis.

Comparing Meta- and Mega-Analytic Single-Gene Results

Among the 90 genes reaching a Bonferroni-corrected level of significance in the covariate-controlled mega-analysis, we performed heterogeneity analysis, which involved per-study linear modeling with inclusion of fixed effects (age, sex, diagnostic status); the beta coefficient and error estimates associated with diagnostic status were combined across studies using inverse variance weighted meta-analysis and forest plots were produced to explore each study's contribution. Generally, the set of meta-analyzed test statistics tracked closely with those of the mega-analysis; for 87 of 90 genes, the estimates of the mega-analyzed β fell within the meta-analytic 95% confidence intervals (Supplementary Table 4 and Supplementary Figure 1). Notably, only thirty-nine of the 90 genes reached a Bonferroni-corrected p-value (for 21,968 genes) under meta-analytic modeling (Supplementary Table 4). Meta-analysis revealed heterogeneity among 26 of these 90 genes (uncorrected Q statistic p-value < 0.05), but the sizes of these effects were very small ($I^2 < 0.01$) compared to chance variability between studies (Supplementary Table 4). The forest plots for the five most significant transcripts (based on covariate-controlled mega-analysis) are shown in Supplementary Figure 1, with the full set of 90 transcripts available for download in the online Supplementary Materials. While many of these 90 transcripts showed consistent direction and magnitude of ASD-related effects across most or all of the study samples (*e.g. LGAL3BP* or *KRLB1*), there was a sizable minority of transcripts for which the effects appeared to be mediated by fewer well-powered studies (*e.g. RPL21P2*).

Sensitivity Analysis of Single-Gene Results

Sensitivity analysis was performed by iteratively leaving out each sample and repeating the mega-analysis (*i.e.*, Jackknifing). We examined the p-values obtained from covariate-controlled mixed-effect models for the 90 transcripts reaching Bonferroni-corrected significance in the full mega-analysis (Supplementary Table 5). Our results indicated that the largest studies (particularly the CHARGE Study and Glatt *et al.,* 2012) may have contributed critically to the mega-analytic effects observed for twelve transcripts (*C14orf28, FAUP1, HABP4, IGF2BP3, MAGOHB, MIR3198-2, MIR7112, MS4A2, NR3C2, RPL21P2,* *RPL24P4, SNORD65*); these transcripts no longer reached nominal significance when the CHARGE and Glatt et al., 2012 samples were withheld. More generally, the sensitivity analysis indicated that the removal of any data set substantially reduced the ability to detect differentially expressed transcripts at a threshold reflecting Bonferroni-correction for the total number of transcripts tested.

Comparison of Single-Gene Findings with Previously Published Meta-Analyses

For all 1,572 genes reaching an FDR q-value < 0.05 in the full covariate-controlled mega-analysis, we also cross-referenced our results with those of recent meta-analyses (Ch'ng et al., 2015; Ning et al., 2015); this comparison is shown in Supplementary Table 3. We found three genes in common with Ning *et al.*; both studies reported under-expression of *OR51F1* and *THOC6* in ASD samples. For *LMBR1L*, we observed over-expression, whereas Ning *et al.* reported under-expression in ASD samples. More generally, Ning *et al.* published only the 20 most significant over- and 20 most significant under-expressed genes, precluding a comprehensive comparison. The lack of replication and discrepancy could be related to differences between studies. Ning *et al.,* combined gene-level summary statistics from multiple sample types (*ex-vivo* blood, transformed lymphoblastoid cell lines (LCL), and *post-mortem* brain). We also cross-referenced our results against a total of six dysregulated gene lists reported by Ch'ng et al. Among their over-expressed genes (as observed in LCL samples, non-LCL blood samples, and all samples combined), we also observed 28 of these genes to reach FDR $q < 0.05$ in our sample, and 25 were dysregulated in the same direction (Supplementary Table 3). Among their under-expressed genes, we observed ten of the same genes to be dysregulated in our sample; 8 showed the same direction of effect (Supplementary Table 3). Among our most significant findings which showed evidence of replication in previous meta-analyses was ASD-related over-expression of *YES1*. This proto-oncogene encodes a non-receptor tyrosine-protein kinase belonging to the src family; stimulation by receptor tyrosine kinases including EGRF, PDGFR, CSF1R, and FGFR leads to recruitment of YES1 protein to the phosphorylated receptor, and activation and phosphorylation of downstream substrates. It is involved in the regulation of cell growth and survival, apoptosis, cell-cell adhesion, cytoskeleton remodeling, and differentiation.

ASD-Associated Gene-Sets

Next, we examined whether functionally related sets of transcripts showed evidence of average differential expression as a group, compared to randomly selected gene-sets of equal size, using the covariate-controlled mega-analytic single-gene results. Gene-set analysis identified 716 sets (among 9249 examined) with at least one significant test hypothesis (Bonferroni $p < 0.05$; depicted in Figure 1, full details in Supplementary Table 6). One hundred and fifty-eight gene-sets were over-expressed; these implicated functions including aerobic energy production, protein metabolism (ribosomal translation, 3′-UTR regulation of translation, nonsense mediated decay at exonic boundaries), nucleotide metabolism (virus-related transcription and life cycle, DNA repair), and genes shown to be over-expressed by up-regulated by MTOR inhibition and ERBB2 over-expression. The majority of gene-sets (525) were under-expressed among ASD cases; these implicated functions involved in innate and adaptive immunity (type I and II interferons, targets of IRSE, NOD-like receptor signaling, RIG1-MDA5-mediated induction if interferon, toll-like receptors 3 and 4, IL-1, IL-6 *via* STAT3, IL-8 biosynthesis, TNF-α *via* NF-κB, B-cell receptor signaling, antigen presentation, and complement cascade), cell survival and signal integration (PI3K-AKT-mTOR, KRAS, MAPK, ERK1 and 2), cell proliferation (G1-to-S transition, G2-M checkpoint, apoptosis, E2F Targets, mitosis), proteolysis, olfactory signal transduction, trophic signaling targets (EGFR, PDGFR), and other transcriptional targets.

We repeated these analyses using the diagnostic group mean differences for the 14,292 gene transcripts obtained using SVA-residualized data. Among a total 7,365 gene-sets evaluated, 287 demonstrated significant effects after Bonferroni-correction for the number of gene-sets tested within each database (Supplementary Table 7). Among these, thirty-one showed evidence of over-expression among ASD cases, including sets reflecting ribosomal functions (translation and 3′UTR regulation of translation), influenza-life cycle and translation, and nonsense mediated decay enhanced by the exon junction. Additionally, 247 showed evidence of under-expression among ASD cases, including sets reflecting immunologic functions (type I and II interferon responses, RIG1-MDA5 signaling, double-stranded RNA-binding, ISRE-regulated transcripts, and TNF-α signaling via NF-kB), cell division (cell cycle, G2-M checkpoint, RB pathway, mitotic spindle, M-G1 transition, S-phase), cell signal integration (mTORC1), targets of transcription regulators (MYC, E2F), and genes involved in the response to DNA damage. Notably, the effects pertaining to members of the RAS-MAPK-ERK signaling axis only emerged when less rigorous correction for family-wise testing was considered (Benjamini-Hochberg $p < 0.05$; Supplementary Table 8). Thus, many of the emergent functions reported in the covariate-controlled analysis showed evidence of replication within the SVA-residualized data.

ASD-Associated Gene Co-expression Network Modules

Network co-expression analyses were performed using non-residualized Z-scaled expression data for genes consistently reported across all studies (genes = 14,931). Plots displaying soft threshold selection criteria are provided in Supplementary Figure 2. We first sought to examine whether the global correlation structures found in the expression data were similar between ASD and controls. Network preservation analysis indicated that modules identified in comparison subjects were all strongly preserved (*Z-summary* > 10) in ASD samples, and the same was true for the reverse analysis (Supplementary Figure 3A). We then sought to fit co-expression networks within the full sample; block-wise gene clustering dendrograms are provided in Supplementary Figure 2. We sought to examine whether each module's expression (summarized as module eigengene values) differed based on diagnostic status. Thirty network modules were identified within the full sample of cases and comparison subjects, and nine showed nominal associations with diagnostic status based on linear mixed models with stepwise removal of covariate effects; we further characterized five of these modules that met a corrected threshold of significance (BH $p < 0.05$) in terms of significantly enriched function, cell-type, and GWAS hits (depicted in Figure 2, with detailed functional and cell-type enrichment results in Supplementary Table 9). Because an unsigned network analysis allows for genes to load onto module eigengenes with both positive and negative correlation coefficients, we also report the number of positive and negative loadings for each module of interest (shown in Figure 2); all but two genes (Grey60 module) loaded with positive signs. Additionally, for each module of interest, the most highly inter-correlated genes are depicted as nodes (within Figure 2), and information regarding their ASD-related differential expression (taken from the covariate-controlled single-gene analysis) is depicted with color coding in the top-right quadrant of each node. Briefly, the sky blue module (73 genes) was over-expressed in ASD and was significantly enriched with NK-cell-related functions (among others) and GWAS signal (seven genes; permuted $p < 0.04$). The midnight blue module (130 genes) was also over-expressed in ASD and was significantly enriched with annotations related to ribosomal translation, viral life cycle, and transcription factor targets, as well as six GWAS hits (permuted $p = 0.90$, *n.s.*). The tan module (163 genes) was under-expressed in ASD and was enriched with annotations related to cell cycle progression, microtubules, mTOR signaling, estrogen responses, MHC class II-mediated antigen presentation, and targets of transcription factors; it also contained 12 genes with GWAS signal (permuted $p = 0.07$). The green-yellow module (181 genes) was under-expressed in ASD and enriched with interferon-α and -γ, RIG1-MDA, and cytosolic DNA sensor signaling, as well as targets of many transcription factors and 12 GWAS hits (permuted $p = 0.60$, *n.s.*). The grey60 module (123 genes) was also under-expressed in ASD and enriched with annotations related to complement, interferon-γ, NOD-like receptor signaling, mTORC1 signaling, as well as several transcription factor targets, markers of the CD14+ monocyte lineage, and five GWAS hits (permuted $p = 0.98$, *n.s.*). The full list of functional annotations for ASD associated modules are provided in Supplementary Table 9. Additionally, for each of the 5 modules of interest, we plotted individual-level module eigengene values for the two diagnostic groups and we created heatmaps depicting gene expression values for each module (Supplementary Figure 4A through E).

Ingenuity Pathway Analysis of Dysregulated Genes and Networks

From the full covariate-controlled mega-analysis, we examined three gene lists of interest (dysregulated at FDR-$q < 0.05$, dysregulated at Bonferroni $p < 0.05$, and genes included in an ASD-associated network module at Benjamini-Hochberg $p < 0.05$). We interrogated these gene-lists using Ingenuity Pathway Analysis software's Core Analysis under default settings and we focused on Canonical Pathways showing strong enrichment across all three lists of interest (Supplementary Table 10). For each selected pathway (EIF2 Signaling, Regulation of eIF4 and p70S6K Signaling, mTOR Signaling, and Interferon Signaling), we isolated the genes and overlaid information on diagnostic group expression differences in order to create color gradients, with red representing over-expression and green representing under-expression in ASD (Supplementary Figure 5). Genes were automatically organized and depicted as mechanistic hypotheses based on IPA's curated knowledgebase. Finally, the Molecule Activity Predictor (MAP) tool was used to make an inference on whether pathway-related emergent cell behavioral functions would be increased or reduced, based on the observed differences in transcript expression. This analysis suggested that protein translation would be increased (based on Regulation of eIF4 and p70S6K Signaling and mTOR Pathway; Supplementary Figure 5, Panels A and B, respectively), and more specifically that translation elongation would be increased, while initiation might be diminished (EIF2 Signaling; Supplementary Figure 5, Panel C). Furthermore, the analysis indicated that actin organization might be diminished (mTOR Pathway; Supplementary Figure 5, Panel B). Examination of the Interferon Signaling Pathway revealed reduced expression of intracellular mediators (*STAT1, STAT2, IRF9*), as well as the transcripts whose expression would typically be increased by signaling through this cascade (Supplementary Figure 5, Panel D).

Blood-Based Transcriptomic Classification

We examined the best performing classification ensemble for each machine type. All machine types performed with relatively comparable accuracy in the training and withheld test samples (Table 2). Artificial neural networks achieved moderate accuracy in the withheld test sample (ROC AUC = 0.69 ± 0.02, sensitivity = 0.65 ± 0.02, specificity = 0.66 ± 0.04). Linear kernel support vector machines performed with similar accuracy (ROC AUC = 0.69 ± 0.02, sensitivity = 0.65 ± 0.05, specificity = 0.66 ± 0.05). Random forests performed with slightly lower accuracy (ROC AUC = 0.67 ± 0.03, sensitivity = 0.63 ± 0.04, specificity = 0.66 ± 0.05). In order to understand which genes facilitated diagnostic discrimination in the selected machines, we examined the frequency with which each gene appeared in the best performing ensembles (Supplementary Table 11A through C); as expected, these gene lists were highly enriched with those identified as differentially-expressed in the single-gene analysis.

Sex-Specific Analyses and Cross-Sex Comparisons

Single-gene analysis—When data were analyzed separately for females ($n_{\text{ASD}} = 90$, $n_{\text{comparison}} = 117$), 19,508 covariate-controlled mixed-effect models produced valid results and 1,609 genes showed at least nominal evidence for dysregulation (uncorrected $p < 0.05$), with 802 over-expressed and 807 under-expressed in ASD, and none reached a corrected level of significance (FDR $q < 0.05$). Among males ($n_{\text{ASD}} = 536$, $n_{\text{comparison}} = 330$), 17,268 gene-level models produced valid statistical results and 2939 showed at least nominal evidence for dysregulation (uncorrected $p < 0.05$), with 1338 over-expressed and 1601 under-expressed in ASD, and 594 reached a corrected level of significance (FDR $q < 0.05$). When examining the intersection of nominally dysregulated gene lists (Supplementary Figure 6; center of Venn diagram), and accounting for the common background of genes analyzed, we observed significant cross-sex overlap of 489 dysregulated (hypergeometric $p < 5.3\times10^{-61}$), 233 over-expressed ($p < 1.4\times10^{-76}$) and 240 under-expressed genes ($p < 4.7\times10^{-64}$). Interestingly, we observed a few genes showing apparently sex-discordant ASD-related effects (Supplementary Figure 6). Also notably, when we compared the absolute values of the differences of least squared means among the top 1% of dysregulated genes for each sex, we observed larger magnitude ASD-related mean differences among females (mean $|\text{difference}_{\text{f}}| = 0.48 \pm 0.06$, mean $|\text{difference}_{\text{m}}| = 0.30 \pm 0.03$; $p < 6.4\times10^{-73}$) and larger F-values among males (mean $F_{\text{f}} = 10.9 \pm 2.4$, mean $F_{\text{m}} = 18.4 \pm 4.6$; $p < 4.2\times10^{-60}$); these differences could potentially be explained by differences and sample size and relative statistical power. Full lists of sex-specific summary statistics will be provided upon request to the corresponding author.

Gene-set analysis—Gene-set analysis was performed using the single-gene statistics derived from covariate-controlled mega-analysis within each sex (Supplementary Figure 6; side panels). Among the 7,350 sets assessed for females, 1,397 reached a liberal threshold of statistical significance (family-wise BH correction) and 368 reached a conservative threshold (family-wise Bonferroni correction). Among males, 7,358 gene-sets were assessed and 817 and 243 met liberal and conservative thresholds, respectively. We compared the results across sexes with respect to the common background of interrogated gene-sets, in order to identify transcriptomic effects that appear to be conserved across sexes; expectedly, many effects observed in the main analysis showed evidence for preservation in both females and males (Supplementary Figure 6; center of Venn diagram). We also highlight gene-sets that reached a conservative threshold for directional dysregulation in one sex, but that showed no evidence for dysregulation (either conservative or liberally-defined) in the opposite sex (Supplementary Figure 6; side panels); we further verified that no biologically similar (or semantically interchangeable) terms show evidence for dysregulation in the opposite sex. Potentially female-specific findings included over-expression of genes involved in DNA repair and under-expression of genes involved in heme metabolism, IL-2 signaling *via* STAT5, IL-15 signaling, transcriptional targets of MEK and miRNA species, and several chromosomal cytobands (Supplementary Figure 6, left panel). Potentially male-specific findings included over-expression of genes involved in O-linked glycosylation and transcriptional targets of MEK (Supplementary Figure 6, right panel). Full lists of sex-specific gene-set summary statistics will be provided upon request to the corresponding author.

Gene co-expression network analysis—For each sex, we repeated network-level analyses using non-residualized Z-scaled expression data. Among females ($n_{\text{ASD}} = 90$, $n_{\text{comparison}} = 117$), modules were bi-directionally highly preserved between ASD affected and comparison subjects (two modules showed moderate preservation; Supplementary Figure 3B). When female cases and comparison subjects were analyzed together, we identified forty-one network modules; linear mixed models identified 8 as nominally associated with ASD, though none survived BH correction for multiple testing (Supplementary Figure 6, left side). Among males ($n_{\text{ASD}}= 536$, $n_{\text{comparison}} = 330$), modules are also bi-directionally highly preserved (Supplementary Figure 3C). When ASD affected and comparison subjects were analyzed together, we identified thirty-two modules, 8 of which were nominally associated with ASD, and 4 survived BH correction for multiple testing (Supplementary Figure 7, right side). On the whole, ASD-associated modules among females (composed of 1236 genes) shared a significant number of genes (299, hypergeometric $p < 6.0\times10^{-75}$) with those implicated among males (composed of 1208 genes). Two ASD-associated modules among females (dark turquoise and midnight blue) did not show greater-than-chance over-representation of genes that were implicated in either the single-gene differential expression analysis or network analysis among males (Supplementary Figure 7, pink bracket). The dark turquoise module was diminished in affected females and was enriched with erythroblast markers and genes involved in heme metabolism, among other functions. The midnight blue module was enhanced in affected females and was enriched with CD4+ and CD+ T-cell-related functions, estrogen-responsive genes, cell proliferation and WNT-beta-catenin signaling. Among males, four modules that were diminished in ASD cases (cyan, dark turquoise, midnight blue, dark turquoise, and yellow) did not show greater-than-chance over-representation of genes that were implicated among female samples (Supplementary Figure 6, blue bracket). Notably, the midnight blue module was enriched with genes involved in neurotrophic signaling, cell adhesion, and B-cell-related functions, and showed a trend toward enrichment with GWAS signal (permuted

$p < .09$). The dark turquoise module was also enriched with B-cell related functions and transcriptional targets of NKX2. The cyan module was enriched with genes involved in cell cycle, estrogen responses, MHCII antigen presentation, and several transcriptional targets, as well as GWAS signal (permuted $p < 0.04$). The large yellow module was enriched with genes involved in innate immune signaling functions, neurodevelopmental processes, neurotrophic signaling, glycerophospholipid metabolism, and numerous transcriptional targets. Full lists of sex-specific summary statistics and module enrichments will be provided upon request to the corresponding author.

DISCUSSION

Our study provides the best-available determination of dysregulated transcripts and characterization of their emergent functions in *ex-vivo* blood samples obtained from individuals with ASD and unrelated comparison subjects. Among the 90 genes reaching a Bonferroni-corrected level of significance based on the full mega-analysis, we found that test statistics tracked closely with those produced by our own corresponding meta-analysis. The top genes and emergent functions also showed similar patterns of dysregulation when an alternate strategy (surrogate variable analysis) was employed. The present findings help to clarify previously published meta-analyses by focusing on *ex-vivo* blood samples, and it extends upon previous studies by testing for directional dysregulation of functional gene-sets and co-expression modules. At the level of emergent function, many of the cell-types and signaling cascades implicated by our work have been associated with ASD previously; we provide an expanded review of these findings in the Supplementary Materials. Here, we focus on findings that have not been extensively discussed in previous work, but we feel may be poised to have a high impact on subsequent ASD research.

Despite evidence suggesting ASD-related increases in the abundance of certain leukocyte subclasses (T/NK-cells), the predominant signature observed in the ASD blood transcriptome was characterized by reduced expression of transcripts subserving innate immune and inflammatory signaling. Among these, type I (α/β) and type II (γ) interferon (IFN)-stimulated signaling cascades were strongly implicated in nearly every level of our analysis and were further supported by secondary re-analysis using Ingenuity Pathway Analysis software. We observed diminished expression of genes and gene expression networks subserving these functions in the blood of ASD cases. IFNs are small glycoprotein cytokines whose functions are primarily understood in the context of mounting immunological responses to infectious agents; both classes exert pleiotropic effects depending on the immunologic milieu, but the type I (INF-α/β) class are best understood as playing a central role in anti-viral responses (Theofilopoulos et al., 2005) and type II (IFN-γ) are associated with inflammatory signaling and responding to a wider variety of intracellular microbes (Boehm et al., 1997; Schroder et al., 2004). The observation of reduced IFN-related gene expression in blood is surprising, particularly in the context of a recent meta-analysis which indicated higher levels of IFN-γ protein in ASD peripheral blood analytes (Masi et al., 2015); three studies found no difference in IFN-α protein levels. These data could be consistent with a decoupling or counter-regulation of IFN signaling in peripheral leukocytes, such that a relatively higher concentration of circulating IFN is eliciting a relatively weaker leukocyte transcriptional response in ASD-affected individuals as compared with typically developing individuals. This interpretation might also be consistent with a small body of literature indicating a higher abundance of less effective NK cell in ASD (see Supplementary Materials for a review).

Notably, a relatively large brain-based RNA sequencing study implicated increased type I and II IFN-related signaling in ASD (Gupta et al., 2014), and a protein-based study of frontal cortex found higher levels of IFN-γ (Li et al., 2009). The present study robustly implicates a peripheral blood gene co-expression module featuring coordinate under-expression of IFN-α/γ-regulated genes; within the same co-expression model, we find enrichment for several signaling cascades that play well-established roles in brain development and synaptic plasticity (*e.g.*, AKT-mTOR, RAS-MEK-MAPK, and WNT signaling pathways; Gkogkas et al., 2013; Crino, 2011; Hoeffer and Klann, 2010; Brambilla et al., 1997; Budnik and Salinas, 2011; Oliva et al., 2013; Stornetta and Zhu, 2011; Sanchez-Ortiz et al., 2014; Yang et al., 2013). Genetic mutations producing hyper-activity of these signaling cascades (*i.e.*, mTOR and RAS) are well-known causes of neurodevelopmental syndromes comorbid with ASD (Hoeffer and Klann, 2010; Alfieri et al., 2014; Crino, 2011; Stornetta and Zhu, 2011; Kelleher and Bear, 2008), with suggestive evidence implicating others (Stornetta and Zhu, 2011; Kalkman, 2012; Levy et al., 2011). One previous study of cortical tissue from idiopathic ASD found reduced expression and activation of mTOR signaling pathway constituents (Nicolini et al., 2015), though less is known about RAS-MAPK signaling axes in the idiopathic ASD brain. While we are not aware of established mechanisms linking IFNs to these pathways, we highlight another recent informatics study that identified inflammatory signaling integrators (NF-κB, JNK, MAPK, TNFs, TGFB1, and MYC) as central to ASD candidate-gene expression networks (Ziats and Rennert, 2011). We also highlight evidence from animal models that IFN-γ plays a role in regulating neuronal development (Li et al., 2010), and when introduced exogenously or up-regulated genetically, can alter synaptic plasticity and development of the visual system, hippocampus and cerebellum (Kim et al., 2002; Vikman et al., 2001; Brask et al., 2004; Ahn et al., 2015; Wang et al., 2004; Maher et al., 2006; Li et al., 2010; Barish et al., 1991). While several mechanisms are now being uncovered, we also suggest that IFN- and other cytokine-induced neuronal MHC class I expression could play a direct role in altering plasticity and brain development (Shatz, 2009; Victório et al., 2012; Gu et al., 2013; Chacon and Boulanger, 2013; Li et al., 2010).

In reviewing our findings in the context of previous work (as detailed within the Supplementary Materials), we must consider a few possibilities: (1) ASD-related transcriptomic signals in leukocytes may be different than those of ASD in the brain (*e.g.,* aerobic metabolism/mitochondrial genes, TNF-α *via* NF-kB, IFN-γ, and others). (2) ASD-related transcriptomic signals in leukocytes may in fact be the opposite of what is seen at the level of circulating protein signal mediators (*e.g.,* IFN-γ, IL-6, and IL-8 signaling). (3) Many of the cell surface receptors whose signaling was down-regulated in our analysis are known to exert their cellular and transcriptional effects through second messengers that are also down-regulated among our results (*e.g.*, EGFR signaling *via* MAPK and PI3K-AKT). To more effectively convey these discrepancies we have attempted to collate our results with the findings of previous transcriptional and protein-based studies in human blood and brain tissues (Figure 3). In reviewing the literature, we noticed that *post-mortem* brain studies have generally shown over-expression of transcripts subserving immunologic functions (Voineagu et al., 2011; Gupta et al., 2014), whereas our results clearly indicate the opposite pattern (reduced expression) in ASD blood samples. This observation could possibly be influenced by differences in the tissue localization of the originating inflammatory insult (*e.g.,* if the inflammatory response originated in the brain), by *post-mortem* hypoxic damage to brain tissues, or by tissue-related differences in negative feedback response (*e.g.*, if inflammation shows more pronounced transcriptional counter-regulation in leukocytes). Additionally, differences in the age of subjects (who tend to be older in *post-mortem* brain studies) or medication usage could contribute to these differences. With respect to the discrepancy between blood-based studies of signaling protein and transcriptional cascade, several phenomena could explain this pattern. One hypothesis might be that long-term over-activity (at the level of circulating protein or intracellular protein expression/phosphorylation/signaling activity), perhaps through feedback mechanisms, is accompanied by down-regulation of the transcripts coding for genes that subserve these signaling functions and of transcripts that are typically induced by these signaling cascades. A related hypothesis might be that ASD is characterized by decoupling between some circulating signaling molecules and their intracellular transcriptional effects. Future mechanistic and within-subject, cross-tissue studies could help shed light on these apparent discrepancies. A full discussion of the findings in our literature search is provided in the Supplementary Materials.

The scientific community has long recognized the concept of an adaptive immune system, classically including antigen-specific T-cells, B-cells, and circulating antibodies. However, we suggest that an expanded concept of immunoplasticity could also include other leukocyte subtype populations and the signaling set-points and feedback mechanisms that govern cell-intrinsic responsivity, as well as the paracrine and hormonal signaling mechanisms that govern interactions between classes of leukocytes and between leukocytes and non-leukocyte cells and tissues (MacGillivray and Kollmann, 2014). An emerging body of literature supports the idea that internal and external environmental factors (*e.g.*, infection and toxicants) exert a persistent influence on later-life immunological milieu (Brodin et al., 2015; Winans et al., 2011; Hansbro et al., 2014; Gbédandé et al., 2013; R. Dietert and Judith T. Zelikoff, 2015; Chen et al., 2011; Luan et al., 2015; Garay et al., 2013). With respect to the present study's findings, we highlight evidence from the fields of immunology and infectious disease, indicating that down-regulation of IFN-γ-responsive genes is a well-characterized phenomenon in chronic infections (Taylor and Mossman, 2013; Kim et al., 2007). A growing literature of animal studies indicates that prenatal immune activation can cause behavioral and neurobiological phenotypes reminiscent of human disorders; as these models begin to reveal the roles of specific signaling molecules and their interactors in typical brain development, we will have a context for understanding how persistent dysregulation of these systems in human subjects might be poised to influence neurodevelopmental phenotypes like ASD. It is also important to recognize that relatively few individuals exposed to environmental risk factors develop ASD-like phenotypes; genetic vulnerabilities impacting immunologic and neurobiologic systems may help explain why only a minority are profoundly affected (Hsiao et al., 2012).

We must recognize a number of limitations pertaining to this study. Our findings are strictly correlational, and as such we cannot infer whether these findings are related to the pathophysiology of ASD or arise in response to internal and external environmental factors or reflect epiphenomena of the syndrome or its treatment. We relied upon the diagnoses and exclusion criteria of the original study authors. We included subjects that would meet the DSM-IV criteria for autistic disorder, as well as those meeting criteria for PPD-NOS and Asperger's Disorder. Thus, it is likely that our results are influenced by both phenotypic and genetic/biological heterogeneity among affected cases, which may diminish power. While we attempted to statistically control for effects related to the study site and demographic covariates shared across studies, a number of unmeasured factors could have influenced the results (*e.g.,* suggestive evidence of stratification amongst leukocyte subclasses), though we attempted to address this concern through replication with the SVA-corrected data, which produced similar results. Another limitation of the present study was its use of an unsigned co-expression network construction approach, which allows for both positively and negatively correlated genes to load onto a given module eigengene; this approach might allow for negatively correlated transcriptomic processes to be grouped together in the same module (so long as they are highly correlated), but this feature can make it more challenging to interpret the meaning of ASD-related module eigengene differences. We found that nearly all genes contained within ASD-associated modules loaded with a positive sign (670 out of 672 genes) and that an overlay of the single-gene differential expression information further facilitated directional interpretation of our unsigned network approach (Figure 2). However, in addition being more easily interpretable, some studies have suggested that signed networks are more likely to be enriched with protein-protein interaction partners and functional pathway relationships than unsigned networks (Ramani et al., 2008; Zhang and Horvath, 2005; Song et al., 2012). Yet another limitation of this study is that the female sample was relatively underpowered for discovery and this may contribute to the impression of sex-differences; future studies should seek to include more female subjects and explicitly examine sex-differences. A power calculation based on the observed characteristics for the top genes observed in the present female sample ($\beta = 0.50$ with *s.d.* = 0.01) suggests that a sample of at least 300 subjects would be sufficiently powered to detect dysregulated transcripts after Bonferroni correction for 20,000 genes ($\alpha = 2.5\times10^{-6}$). Nonetheless, heterogeneity of genomic characteristics among patients could be the norm rather than the exception (Campbell et al., 2013; Diaz-Beltran et al., 2016). Despite these limitations, we feel this work makes a valid and valuable contribution to the transcriptomic characterization of the circulating immunologic milieu in ASD and highlights signaling mechanisms through which immunologic and neurobiological systems could interact.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

D.S.T., J.L.H., and S.J.G. were supported by grants from the U.S. National Institutes of Health, the Gerber Foundation, the Sidney R. Baer, Jr. Foundation, and NARSAD: The Brain & Behavior Research Foundation. D.S.T. was also supported by the Autism Speaks foundation (Weatherstone Pre-Doctoral Training Grant #9645). Microarray and clinical data were provided by the Childhood Autism Risks from Genetics and the Environment

(CHARGE) study (original studies funded by NIEHS R01-ES015359, P01-ES011269, P30 ES023513; NICHD U54 HD079125: U.S. EPA #R-829388 & R833292; and UC Davis MIND Institute). Additionally, CHARGE study investigators B.S.S., I.H.-P. and F.R.S. assisted in preparing data, provided instrumental feedback in planning the analysis, and provided thoughtful revisions to the manuscript. Additionally, microarray and clinical data were provided by S.W.K. and colleagues, who helped to prepare the data, identified duplicated individuals across study samples, assisted in quality control of microarray data and provided thoughtful revisions to the manuscript including the surrogate variable analysis (original studies funded by NIMH 01MH107205 and R01MH085143). S.W.K was supported by NIMH R01MH107205 and P50MH094267. Interpretation of network-level analyses and IPA analyses performed with assistance from Y.Z.-J. and further guidance from S.V.F.. Assistance with the identification and solicitation of eligible study data was supported by R.B. and J.C.. The machine learning classification pipeline was created and maintained by T.P.Q.. We thank all authors of the previous studies that composed our mega-analytic sample, and specifically we thank Mark D. Alter, and Jeffrey P. Gregg, as well as their colleagues, for generously depositing their microarray data in publicly accessible repositories. We gratefully acknowledge all participating individuals and their families. The authors declare that no conflicts of interest influenced this study.

References

Ahn J, Lee J, Kim S. Interferon-gamma inhibits the neuronal differentiation of neural progenitor cells by inhibiting the expression of Neurogenin2 via the JAK/STAT1 pathway. Biochem Biophys Res Commun. 2015; 466:52–9. [PubMed: 26325468]

Alfieri P, Piccini G, Caciolo C, Perrino F, Gambardella ML, Mallardi M, Cesarini L, Leoni C, Leone D, Fossati C, Selicorni A, Digilio MC, Tartaglia M, Mercuri E, Zampino G, Vicari S. Behavioral profile in RASopathies. Am J Med Genet A. 2014; 164A:934–42. [PubMed: 24458522]

Alter MD, Kharkar R, Ramsey KE, Craig DW, Melmed RD, Grebe TA, Bay RC, Ober-Reynolds S, Kirwan J, Jones JJ, Turner JB, Hen R, Stephan DA. Autism and increased paternal age related changes in global levels of gene expression regulation. PLoS One. 2011; 6:e16715. [PubMed: 21379579]

American Psychiatric Association. DSM-V. 2013; 20:31–32. 87–88, 100–104, 155–165.

Barish ME, Mansdorf NB, Raissdana SS. γ-Interferon promotes differentiation of cultured cortical and hippocampal neurons. Dev Biol. 1991; 144:412–423. [PubMed: 1901286]

Becker N, Werft W, Toedt G, Lichter P, Benner A. penalizedSVM: a R-package for feature selection SVM classification. Bioinformatics. 2009; 25:1711–2. [PubMed: 19398451]

Boehm U, Klamp T, Groot M, Howard JC. Cellular responses to interferon-gamma. Annu Rev Immunol. 1997; 15:749–95. [PubMed: 9143706]

Brambilla R, Gnesutta N, Minichiello L, White G, Roylance AJ, Herron CE, Ramsey M, Wolfer DP, Cestari V, Rossi-Arnaud C, Grant SGN, Chapman PF, Lipp HP, Sturani E, Klein R. A role for the Ras signalling pathway in synaptic transmission and long- term memory. Nature. 1997; 390:284–286.

Brask J, Kristensson K, Hill RH. Exposure to interferon-γ during synaptogenesis increases inhibitory activity after a latent period in cultured rat hippocampal neurons. Eur J Neurosci. 2004; 19:3193–3201. [PubMed: 15217375]

Brodin P, Jojic V, Gao T, Bhattacharya S, Angel CJL, Furman D, Shen-Orr S, Dekker CL, Swan GE, Butte AJ, Maecker HT, Davis MM. Variation in the Human Immune System Is Largely Driven by Non-Heritable Influences. Cell. 2015; 160:37–47. [PubMed: 25594173]

Budnik V, Salinas PC. Wnt signaling during synaptic development and plasticity. Curr Opin Neurobiol. 2011; 21:151–9. [PubMed: 21239163]

Campbell MG, Kohane IS, Kong SW. Pathway-based outlier method reveals heterogeneous genomic structure of autism in blood transcriptome. BMC Med Genomics. 2013; 6:34. [PubMed: 24063311]

Carvalho BS, Irizarry Ra. A framework for oligonucleotide microarray preprocessing. Bioinformatics. 2010; 26:2363–2367. [PubMed: 20688976]

Ch'ng C, Kwok W, Rogic S, Pavlidis P. Meta-Analysis of Gene Expression in Autism Spectrum Disorder. Autism Res. 2015; 8:593–608. [PubMed: 25720351]

Chacon MA, Boulanger LM. MHC class I protein is expressed by neurons and neural progenitors in mid-gestation mouse brain. Mol Cell Neurosci. 2013; 52:117–127. [PubMed: 23147111]

Chanda P, Huang H, Arking DE, Bader JS. Fast Association Tests for Genes with FAST. PLoS One. 2013:8.

Chen E, Miller GE, Kobor MS, Cole SW. Maternal warmth buffers the effects of low early-life socioeconomic status on pro-inflammatory signaling in adulthood. Mol Psychiatry. 2011; 16:729–37. [PubMed: 20479762]

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. Age-dependent brain gene expression and copy number anomalies in autism suggest distinct pathological processes at young versus mature ages. PLoS Genet. 2012:8.

Crino PB. mTOR: A pathogenic signaling pathway in developmental brain malformations. Trends Mol Med. 2011; 17:734–42. [PubMed: 21890410]

Diaz-Beltran L, Esteban FJ, Wall DP. A common molecular signature in ASD gene expression: following Root 66 to autism. Transl Psychiatry. 2016; 6:e705. [PubMed: 26731442]

Durinck S, Spellman PT, Birney E, Huber W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nat Protoc. 2009; 4:1184–91. [PubMed: 19617889]

Enstrom AM, Lit L, Onore CE, Gregg JP, Hansen RL, Pessah IN, Hertz-Picciotto I, Van de Water JA, Sharp FR, Ashwood P. Altered gene expression and function of peripheral blood natural killer cells in children with autism. Brain Behav Immun. 2009; 23:124–133. [PubMed: 18762240]

Garay PA, Hsiao EY, Patterson PH, McAllister AK. Maternal immune activation causes age- and region-specific changes in brain cytokines in offspring throughout development. Brain Behav Immun. 2013; 31:54–68. [PubMed: 22841693]

Garbett K, Ebert PJ, Mitchell A, Lintas C, Manzi B, Mirnics K, Persico AM. Immune transcriptome alterations in the temporal cortex of subjects with autism. Neurobiol Dis. 2008; 30:303–311. [PubMed: 18378158]

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. Most genetic risk for autism resides with common variation. Nat Genet. 2014; 46:881–5. [PubMed: 25038753]

Gaujoux R, Seoighe C. CellMix: a comprehensive toolbox for gene expression deconvolution. Bioinformatics. 2013; 29:2211–2. [PubMed: 23825367]

Gautier L, Cope L, Bolstad BM, Irizarry RA. affy--analysis of Affymetrix GeneChip data at the probe level. Bioinformatics. 2004; 20:307–15. [PubMed: 14960456]

Gbédandé K, Varani S, Ibitokou S, Houngbegnon P, Borgella S, Nouatin O, Ezinmegnon S, Adeothy A-L, Cottrell G, Massougbodji A, Moutairou K, Troye-Blomberg M, Deloron P, Fievet N, Luty AJF. Malaria modifies neonatal and early-life toll-like receptor cytokine responses. Infect Immun. 2013; 81:2686–96. [PubMed: 23690399]

Geschwind DH. Autism: many genes, common pathways? Cell. 2008; 135:391–5. [PubMed: 18984147]

Ginsberg MR, Rubin RA, Falcone T, Ting AH, Natowicz MR. Brain transcriptional and epigenetic associations with autism. PLoS One. 2012; 7:e44736. [PubMed: 22984548]

Gkogkas CG, Khoutorsky A, Ran I, Rampakakis E, Nevarko T, Weatherill DB, Vasuta C, Yee S, Truitt M, Dallaire P, Major F, Lasko P, Ruggero D, Nader K, Lacaille J-C, Sonenberg N. Autism-related deficits via dysregulated eIF4E-dependent translational control. Nature. 2013; 493:371–7. [PubMed: 23172145]

Glatt SJ, Tsuang MT, Winn M, Chandler SD, Collins M, Lopez L, Weinfeld M, Carter C, Schork N, Pierce K, Courchesne E. Blood-based gene expression signatures of infants and toddlers with autism. J Am Acad Child Adolesc Psychiatry. 2012; 51:934–44. e2. [PubMed: 22917206]

Gregg JP, Lit L, Baron CA, Hertz-Picciotto I, Walker W, Davis RA, Croen LA, Ozonoff S, Hansen R, Pessah IN, Sharp FR. Gene expression changes in children with autism. Genomics. 2008; 91:22–29. [PubMed: 18006270]

Gu S, Fellerhoff B, Müller N, Laumbacher B, Wank R. Paradoxical downregulation of HLA-A expression by IFNγ associated with schizophrenia and noncoding genes. Immunobiology. 2013; 218:738–44. [PubMed: 23083632]

Gupta S, Ellis SE, Ashar FN, Moes A, Bader JS, Zhan J, West AB, Arking DE. Transcriptome analysis reveals dysregulation of innate immune response genes and neuronal activity-dependent genes in autism. Nat Commun. 2014; 5:5748. [PubMed: 25494366]

Hansbro PM, Starkey MR, Mattes J, Horvat JC. Pulmonary immunity during respiratory infections in early life and the development of severe asthma. Ann Am Thorac Soc. 2014; 11(Suppl 5):S297–302. [PubMed: 25525736]

Hoeffer CA, Klann E. mTOR signaling: at the crossroads of plasticity, memory and disease. Trends Neurosci. 2010; 33:67–75. [PubMed: 19963289]

Hsiao EY, McBride SW, Chow J, Mazmanian SK, Patterson PH. Modeling an autism risk factor in mice leads to permanent immune dysregulation. Proc Natl Acad Sci U S A. 2012; 109:12776–81. [PubMed: 22802640]

Huang H, Chanda P, Alonso A, Bader JS, Arking DE. Gene-Based tests of association. PLoS Genet. 2011:7.

Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics. 2003; 4:249–64. [PubMed: 12925520]

De Jay N, Papillon-Cavanagh S, Olsen C, El-Hachem N, Bontempi G, Haibe-Kains B. mRMRe: an R package for parallelized mRMR ensemble feature selection. Bioinformatics. 2013; 29:2365–8. [PubMed: 23825369]

Johannes M, Fröhlich H, Sültmann H, Beissbarth T. pathClass: an R-package for integration of pathway knowledge into support vector machines for biomarker discovery. Bioinformatics. 2011; 27:1442–3. [PubMed: 21450711]

Kalkman HO. A review of the evidence for the canonical Wnt pathway in autism spectrum disorders. Mol Autism. 2012; 3:10. [PubMed: 23083465]

Kelleher RJ, Bear MF. The autistic neuron: troubled translation? Cell. 2008; 135:401–6. [PubMed: 18984149]

Kim IJ, Beck HN, Lein PJ, Higgins D. Interferon γ Induces Retrograde Dendritic Retraction and Inhibits Synapse Formation. J Neurosci. 2002; 22:4530–4539. [PubMed: 12040060]

Kim SK, Fouts AE, Boothroyd JC. Toxoplasma gondii dysregulates IFN-γ-inducible gene expression in human fibroblasts: Insights from a genome-wide transcriptional profiling. J Immunol. 2007; 178:5154–5165. [PubMed: 17404298]

Kong SW, Collins CD, Shimizu-Motohashi Y, Holm IA, Campbell MG, Lee I-H, 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. Characteristics and predictive value of blood transcriptome signature in males with autism spectrum disorders. PLoS One. 2012; 7:e49475. [PubMed: 23227143]

Kong SW, Shimizu-Motohashi Y, Campbell MG, Lee IH, Collins CD, Brewster SJ, Holm Ia, Rappaport L, Kohane IS, Kunkel LM. Peripheral blood gene expression signature differentiates children with autism from unaffected siblings. Neurogenetics. 2013; 14:143–52. [PubMed: 23625158]

Kuwano Y, Kamio Y, Kawai T, Katsuura S, Inada N, Takaki A, Rokutan K. Autism-associated gene expression in peripheral leucocytes commonly observed between subjects with autism and healthy women having autistic children. PLoS One. 2011:6.

Kuznetsova A, Brockhoff PB, Christensen RHB. lmerTest. 2016

Langfelder P, Luo R, Oldham MC, Horvath S. Is my network module preserved and reproducible? PLoS Comput. Biol. 2011; 7:e1001057.

Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics. 2012; 28:882–3. [PubMed: 22257669]

Leek JT, Storey JD. Capturing heterogeneity in gene expression studies by surrogate variable analysis. PLoS Genet. 2007; 3:1724–35. [PubMed: 17907809]

Levy D, Ronemus M, Yamrom B, Lee Yha, Leotta A, Kendall J, Marks S, Lakshmi B, Pai D, Ye K, Buja A, Krieger A, Yoon S, Troge J, Rodgers L, Iossifov I, Wigler M. Rare De Novo and

Transmitted Copy-Number Variation in Autistic Spectrum Disorders. Neuron. 2011; 70:886–897. [PubMed: 21658582]

Li L, Walker TL, Zhang Y, Mackay EW, Bartlett PF. Endogenous Interferon Directly Regulates Neural Precursors in the Non-Inflammatory Brain. J Neurosci. 2010; 30:9038–9050. [PubMed: 20610738]

Li X, Chauhan A, Sheikh AM, Patil S, Chauhan V, Li X-M, Ji L, Brown T, Malik M. Elevated immune response in the brain of autistic patients. J Neuroimmunol. 2009; 207:111–6. [PubMed: 19157572]

Luan R, Cheng H, Li L, Zhao Q, Liu H, Wu Z, Zhao L, Yang J, Hao J, Yin Z. Maternal Lipopolysaccharide Exposure Promotes Immunological Functional Changes in Adult Offspring CD4+ T Cells. Am J Reprod Immunol. 2015; 73:522–35. [PubMed: 25640465]

MacGillivray DM, Kollmann TR. The role of environmental factors in modulating immune responses in early life. Front Immunol. 2014; 5:434. [PubMed: 25309535]

Maher FO, Clarke RM, Kelly A, Nally RE, Lynch MA. Interaction between interferon gamma and insulin-like growth factor-1 in hippocampus impacts on the ability of rats to sustain long-term potentiation. J Neurochem. 2006; 96:1560–71. [PubMed: 16464236]

Masi A, Quintana DS, Glozier N, Lloyd AR, Hickie IB, Guastella AJ. Cytokine aberrations in autism spectrum disorder: a systematic review and meta-analysis. Mol Psychiatry. 2015; 20:440–6. [PubMed: 24934179]

Meyer D, Dimitriadou E, Hornik K, Weingessel A, Leisch F, Chang C-C, Lin C-C. e1071: Misc Functions of the Department of Statistics, Probability Theory Group. 2015

Mistry M, Gillis J, Pavlidis P. Genome-wide expression profiling of schizophrenia using a large combined cohort. Mol Psychiatry. 2013; 18:215–25. [PubMed: 22212594]

Nicolini C, Ahn Y, Michalski B, Rho JM, Fahnestock M. Decreased mTOR signaling pathway in human idiopathic autism and in rats exposed to valproic acid. Acta Neuropathol Commun. 2015; 3:3. [PubMed: 25627160]

Ning LF, Yu YQ, GuoJi ET, Kou CG, Wu YH, Shi JP, Ai LZ, Yu Q. Meta-analysis of differentially expressed genes in autism based on gene expression data. Genet Mol Res. 2015; 14:2146–55. [PubMed: 25867362]

Oliva CA, Vargas JY, Inestrosa NC. Wnts in adult brain: from synaptic plasticity to cognitive deficiencies. Front Cell Neurosci. 2013; 7:224. [PubMed: 24348327]

Pages H, Marc C, Falcon S, Li N. AnnotationDbi: Annotation Database Interface. 2016

Dietert R, Judith T, Zelikoff R. Pediatric Immune Dysfunction and Health Risks Following Early-Life Immune Insult. Curr Pediatr Rev. 2015:5. [PubMed: 25938379]

Ramani AK, Li Z, Hart GT, Carlson MW, Boutz DR, Marcotte EM. A map of human protein interactions derived from co-expression of human mRNAs and their orthologs. Mol Syst Biol. 2008; 4:180. [PubMed: 18414481]

Ripley B, Venables W. nnet: Feed-Forward Neural Networks and Multinomial Log-Linear Models. 2015

Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015:gkv007.

De Rubeis S, He X, Goldberg AP, Poultney CS, Samocha K, Ercument Cicek A, Kou Y, Liu L, Fromer M, Walker S, Singh T, Klei L, Kosmicki J, Fu S-C, Aleksic B, Biscaldi M, Bolton PF, Brownfeld JM, Cai J, Campbell NG, Carracedo A, Chahrour MH, Chiocchetti AG, Coon H, Crawford EL, Crooks L, Curran SR, Dawson G, Duketis E, Fernandez BA, Gallagher L, Geller E, Guter SJ, Sean Hill R, Ionita-Laza I, Jimenez Gonzalez P, Kilpinen H, Klauck SM, Kolevzon A, Lee I, Lei J, Lehtimäki T, Lin C-F, Ma'ayan A, Marshall CR, McInnes AL, Neale B, Owen MJ, Ozaki N, Parellada M, Parr JR, Purcell S, Puura K, Rajagopalan D, Rehnström K, Reichenberg A, Sabo A, Sachse M, Sanders SJ, Schafer C, Schulte-Rüther M, Skuse D, Stevens C, Szatmari P, Tammimies K, Valladares O, Voran A, Wang L-S, Weiss LA, Jeremy Willsey A, Yu TW, Yuen RKC, 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. Synaptic, transcriptional and chromatin genes disrupted in autism. Nature. 2014; 515:209–15. [PubMed: 25363760]

Sanchez-Ortiz E, Cho W, Nazarenko I, Mo W, Chen J, Parada LF. NF1 regulation of RAS/ERK signaling is required for appropriate granule neuron progenitor expansion and migration in cerebellar development. Genes Dev. 2014; 28:2407–20. [PubMed: 25367036]

Sanders SJ, Ercan-Sencicek AG, Hus V, Luo R, Murtha MT, Moreno-De-Luca D, Chu SH, Moreau MP, Gupta AR, Thomson SA, Mason CE, Bilguvar K, Celestino-Soper PBS, Choi M, Crawford EL, Davis L, Wright NRD, Dhodapkar RM, DiCola M, DiLullo NM, Fernandez TV, Fielding-Singh V, Fishman DO, Frahm S, Garagaloyan R, Goh GS, Kammela S, Klei L, Lowe JK, Lund SC, McGrew AD, Meyer KA, Moffat WJ, Murdoch JD, O'Roak BJ, Ober GT, Pottenger RS, Raubeson MJ, Song Y, Wang Q, Yaspan BL, Yu TW, Yurkiewicz IR, Beaudet AL, Cantor RM, Curland M, Grice DE, Günel M, Lifton RP, Mane SM, Martin DM, Shaw CA, Sheldon M, Tischfield JA, Walsh CA, Morrow EM, Ledbetter DH, Fombonne E, Lord C, Martin CL, Brooks AI, Sutcliffe JS, Cook EH, Geschwind D, Roeder K, Devlin B, State MW. Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. Neuron. 2011; 70:863–85. [PubMed: 21658581]

Sanders SJ, Murtha MT, Gupta AR, Murdoch JD, Raubeson MJ, Willsey 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, Šestan N, Lifton RP, Günel M, Roeder K, Geschwind DH, Devlin B, State MW. De novo mutations revealed by whole-exome sequencing are strongly associated with autism. Nature. 2012; 485:237–241. [PubMed: 22495306]

Sandin S, Lichtenstein P, Kuja-Halkola R, Larsson H, Hultman CM, Reichenberg A. The familial risk of autism. JAMA. 2014; 311:1770–7. [PubMed: 24794370]

Schroder K, Hertzog PJ, Ravasi T, Hume DA. Interferon-gamma: an overview of signals, mechanisms and functions. J Leukoc Biol. 2004; 75:163–89. [PubMed: 14525967]

Schwarzer G. meta. 2016

Seifuddin F, Pirooznia M, Judy JT, Goes FS, Potash JB, Zandi PP. Systematic review of genome-wide gene expression studies of bipolar disorder. BMC Psychiatry. 2013; 13:213. [PubMed: 23945090]

Sekhon JS. Matching: Multivariate and Propensity Score Matching with Balance Optimization. 2015

Shatz CJ. MHC class I: an unexpected role in neuronal plasticity. Neuron. 2009; 64:40–5. [PubMed: 19840547]

Sing T, Sander O, Beerenwinkel N, Lengauer T. ROCR: visualizing classifier performance in R. Bioinformatics. 2005; 21:3940–1. [PubMed: 16096348]

Song L, Langfelder P, Horvath S. Comparison of co-expression measures: mutual information, correlation, and model based indices. BMC Bioinformatics. 2012; 13:328. [PubMed: 23217028]

Stamova B, Green PG, Tian Y, Hertz-Picciotto I, Pessah IN, Hansen R, Yang X, Teng J, Gregg JP, Ashwood P, Van de Water J, Sharp FR. Correlations between gene expression and mercury levels in blood of boys with and without autism. Neurotox Res. 2011; 19:31–48. [PubMed: 19937285]

Stegle O, Parts L, Piipari M, Winn J, Durbin R. Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. Nat Protoc. 2012; 7:500–7. [PubMed: 22343431]

Storey JD. The positive false discovery rate: a Bayesian interpretation and the q-value. Ann Stat. 2003; 31:2013–2035.

Stornetta RL, Zhu JJ. Ras and Rap signaling in synaptic plasticity and mental disorders. Neuroscientist. 2011; 17:54–78. [PubMed: 20431046]

Taylor KE, Mossman KL. Recent advances in understanding viral evasion of type I interferon. Immunology. 2013; 138:190–7. [PubMed: 23173987]

Theofilopoulos AN, Baccala R, Beutler B, Kono DH. Type I interferons (alpha/beta) in immunity and autoimmunity. Annu Rev Immunol. 2005; 23:307–36. [PubMed: 15771573]

Tian Y, Green PG, Stamova B, Hertz-Picciotto I, Pessah IN, Hansen R, Yang X, Gregg JP, Ashwood P, Jickling G, Van de Water J, Sharp FR. Correlations of gene expression with blood lead levels in children with autism compared to typically developing controls. Neurotox Res. 2011; 19:1–13. [PubMed: 19921347]

Tillé Y, Matei A. sampling: Survey Sampling. 2015

Väremo L, Nielsen J, Nookaew I. Enriching the gene set analysis of genome-wide data by incorporating directionality of gene expression and combining statistical hypotheses and methods. Nucleic Acids Res. 2013; 41:4378–91. [PubMed: 23444143]

Victório SCS, Cartarozzi LP, Hell RCR, Oliveira ALR. Decreased MHC I expression in IFN γ mutant mice alters synaptic elimination in the spinal cord after peripheral injury. J Neuroinflammation. 2012; 9:88. [PubMed: 22564895]

Vikman KS, Owe-Larsson B, Brask J, Kristensson KS, Hill RH. Interferon-γ-induced changes in synaptic activity and AMPA receptor clustering in hippocampal cultures. Brain Res. 2001; 896:18–29. [PubMed: 11277968]

Voineagu I, Wang X, Johnston P, Lowe JK, Tian Y, Horvath S, Mill J, Cantor RM, Blencowe BJ, Geschwind DH. Transcriptomic analysis of autistic brain reveals convergent molecular pathology. Nature. 2011; 474:380–4. [PubMed: 21614001]

Wang J, Lin W, Popko B, Campbell IL. Inducible production of interferon-γ in the developing brain causes cerebellar dysplasia with activation of the Sonic hedgehog pathway. Mol Cell Neurosci. 2004; 27:489–496. [PubMed: 15555926]

Watkins NA, Gusnanto A, de Bono B, De S, Miranda-Saavedra D, Hardie DL, Angenent WGJ, Attwood AP, Ellis PD, Erber W, Foad NS, Garner SF, Isacke CM, Jolley J, Koch K, Macaulay IC, Morley SL, Rendon A, Rice KM, Taylor N, Thijssen-Timmer DC, Tijssen MR, van der Schoot CE, Wernisch L, Winzer T, Dudbridge F, Buckley CD, Langford CF, Teichmann S, Göttgens B, Ouwehand WH. A HaemAtlas: characterizing gene expression in differentiated human blood cells. Blood. 2009; 113:e1–9. [PubMed: 19228925]

Winans B, Humble MC, Lawrence BP. Environmental toxicants and the developing immune system: a missing link in the global battle against infectious disease? Reprod. Toxicol. 2011; 31:327–36.

Wu, J., Irizarry, RA. gcrma: Background Adjustment Using Sequence Information. 2016. http://bioconductor.org/packages/2.8/bioc/html/gcr

Yang K, Cao F, Sheikh AM, Malik M, Wen G, Wei H, Ted Brown W, Li X. Up-regulation of Ras/Raf/ERK1/2 signaling impairs cultured neuronal cell migration, neurogenesis, synapse formation, and dendritic spine development. Brain Struct Funct. 2013; 218:669–82. [PubMed: 22555958]

Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. Stat Appl Genet Mol Biol. 2005; 4 Article17.

Ziats MN, Rennert OM. Expression Profiling of Autism Candidate Genes during Human Brain Development Implicates Central Immune Signaling Pathways. PLoS One. 2011; 6:e24691. [PubMed: 21935439]

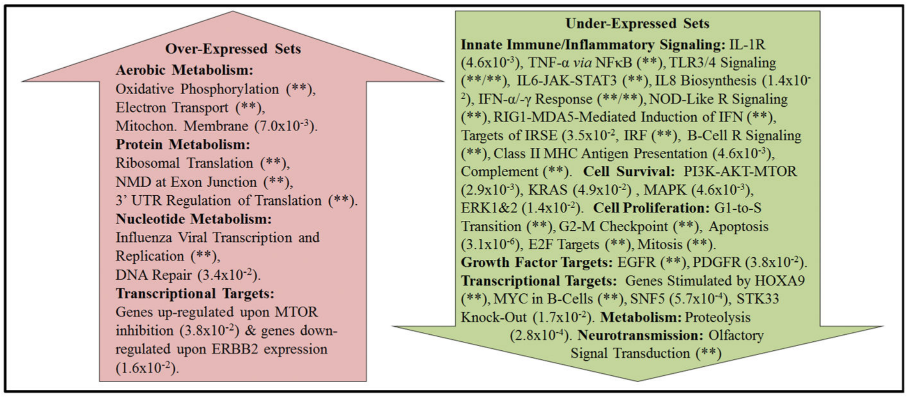
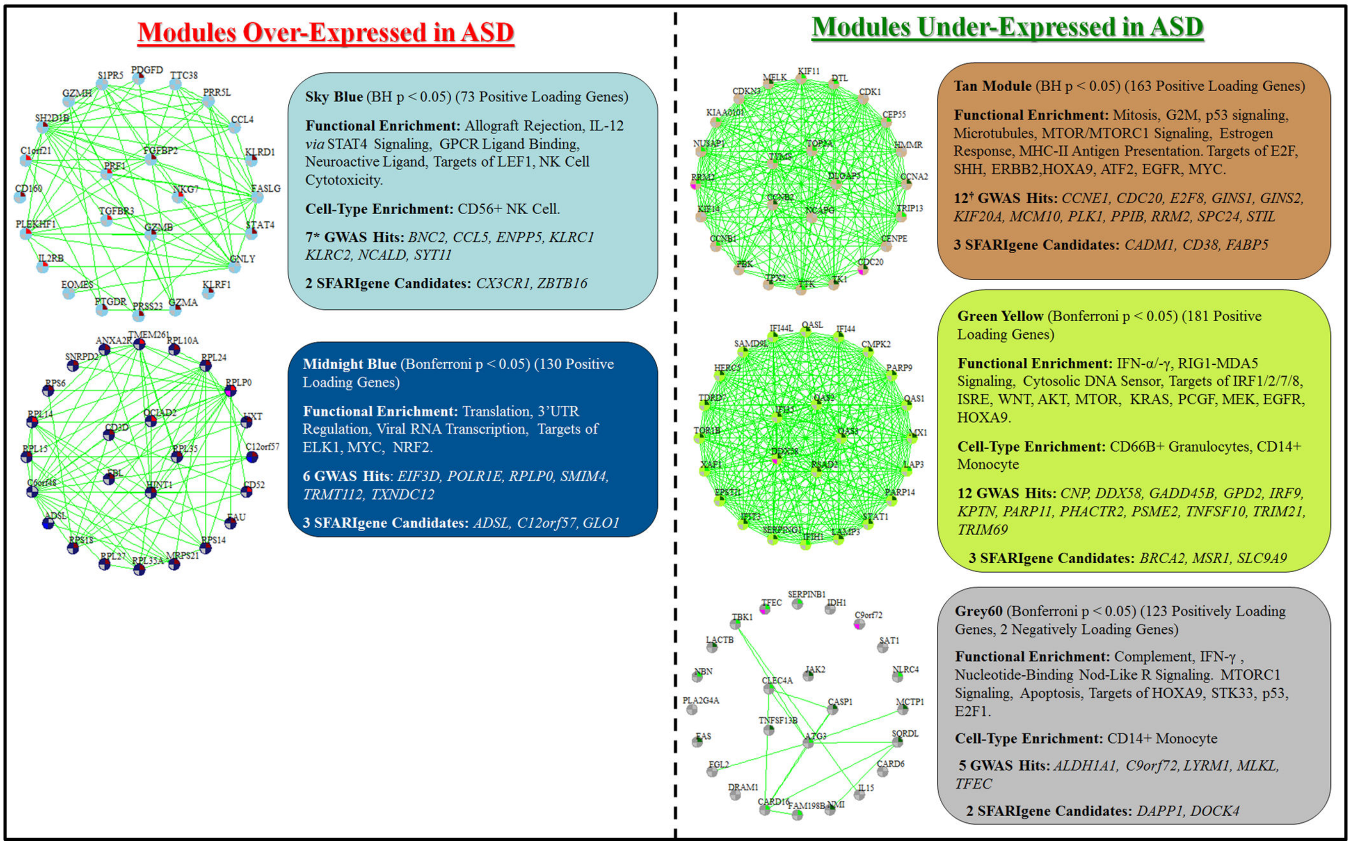

Over-Expressed Sets
Aerobic Metabolism:
Oxidative Phosphorylation (**),
Electron Transport (**),
Mitochon. Membrane (7.0x10^{-3}).
Protein Metabolism:
Ribosomal Translation (**),
NMD at Exon Junction (**),
3’ UTR Regulation of Translation (**).
Nucleotide Metabolism:
Influenza Viral Transcription and Replication (**),
DNA Repair (3.4x10^{-2}).
Transcriptional Targets:
Genes up-regulated upon MTOR inhibition (3.8x10^{-2}) & genes down-regulated upon ERBB2 expression (1.6x10^{-2}).
Under-Expressed Sets
Innate Immune/Inflammatory Signaling: IL-1R (4.6x10^{-3}), TNF-α via NFκB (**), TLR3/4 Signaling (**/**), IL6-JAK-STAT3 (**), IL8 Biosynthesis (1.4x10^{-2}), IFN-α/-γ Response (**/**), NOD-Like R Signaling (**), RIG1-MDA5-Mediated Induction of IFN (**), Targets of IRSE (3.5x10^{-2}, IRF (**), B-Cell R Signaling (**), Class II MHC Antigen Presentation (4.6x10^{-3}), Complement (**).
Cell Survival: PI3K-AKT-MTOR (2.9x10^{-3}), KRAS (4.9x10^{-2}) , MAPK (4.6x10^{-3}), ERK1&2 (1.4x10^{-2}).
Cell Proliferation: G1-to-S Transition (**), G2-M Checkpoint (**), Apoptosis (3.1x10^{-6}), E2F Targets (**), Mitosis (**).
Growth Factor Targets: EGFR (**), PDGFR (3.8x10^{-2}).
Transcriptional Targets: Genes Stimulated by HOXA9 (**), MYC in B-Cells (**), SNF5 (5.7x10^{-4}), STK33 Knock-Out (1.7x10^{-2}).
Metabolism: Proteolysis (2.8x10^{-4}).
Neurotransmission: Olfactory Signal Transduction (**)

Figure 1.

Results of Permutation-Based Gene-Set Analysis.

Test statistics from the covariate-controlled single-gene mega-analysis (differences in diagnostic group means after adjustment for covariate effects) were supplied for permutation-based gene-set analysis. As described in the Methods section, this approach assesses whether a given *a priori*-defined set of genes, on average, shows more evidence of an ASD-associated expression difference than randomly selected gene-sets of equal size. Here we show results that reaches a Bonferroni-corrected $p < 0.05$. The functional themes of over-expressed gene-sets are shown within the red-colored upward-pointing arrow, and the functional themes of under-expressed gene-sets are shown within the green-colored downward-pointing arrow. *P*-values (Bonferroni-corrected for the number of sets per database) are displayed in parenthesis, and *p*-values $< 1\times10^{-6}$ (reflecting the minimum possible *p*-value based on the number of permutations) are denoted with **. Full results of this analysis are available in Supplementary Table 6 and gene-set names can be examined within the Molecular Signature Database for additional context.



Sky Blue (BH p < 0.05) (73 Positive Loading Genes)

Functional Enrichment: Allograft Rejection, IL-12 *via* STAT4 Signaling, GPCR Ligand Binding, Neuroactive Ligand, Targets of LEF1, NK Cell Cytotoxicity.

Cell-Type Enrichment: CD56+ NK Cell.

7* GWAS Hits: *BNC2, CCL5, ENPP5, KLRC1 KLRC2, NCALD, SYT11*

2 SFARIgene Candidates: *CX3CR1, ZBTB16*

Midnight Blue (Bonferroni p < 0.05) (130 Positive Loading Genes)

Functional Enrichment: Translation, 3'UTR Regulation, Viral RNA Transcription, Targets of ELK1, MYC, NRF2.

6 GWAS Hits: *EIF3D, POLR1E, RPLP0, SMIM4, TRMT112, TXNDC12*

3 SFARIgene Candidates: *ADSL, C12orf57, GLO1*

Tan Module (BH p < 0.05) (163 Positive Loading Genes)

Functional Enrichment: Mitosis, G2M, p53 signaling, Microtubules, MTOR/MTORC1 Signaling, Estrogen Response, MHC-II Antigen Presentation. Targets of E2F, SHH, ERBB2,HOXA9, ATF2, EGFR, MYC.

12† GWAS Hits: *CCNE1, CDC20, E2F8, GINS1, GINS2, KIF20A, MCM10, PLK1, PPIB, RRM2, SPC24, STIL*

3 SFARIgene Candidates: *CADM1, CD38, FABP5*

Green Yellow (Bonferroni p < 0.05) (181 Positive Loading Genes)

Functional Enrichment: IFN-α/-γ, RIG1-MDA5 Signaling, Cytosolic DNA Sensor, Targets of IRF1/2/7/8, ISRE, WNT, AKT, MTOR, KRAS, PCGF, MEK, EGFR, HOXA9.

Cell-Type Enrichment: CD66B+ Granulocytes, CD14+ Monocyte

12 GWAS Hits: *CNP, DDX58, GADD45B, GPD2, IRF9, KPTN, PARP11, PHACTR2, PSME2, TNFSF10, TRIM21, TRIM69*

3 SFARIgene Candidates: *BRCA2, MSR1, SLC9A9*

Grey60 (Bonferroni p < 0.05) (123 Positively Loading Genes, 2 Negatively Loading Genes)

Functional Enrichment: Complement, IFN-γ , Nucleotide-Binding Nod-Like R Signaling. MTORC1 Signaling, Apoptosis, Targets of HOXA9, STK33, p53, E2F1.

Cell-Type Enrichment: CD14+ Monocyte

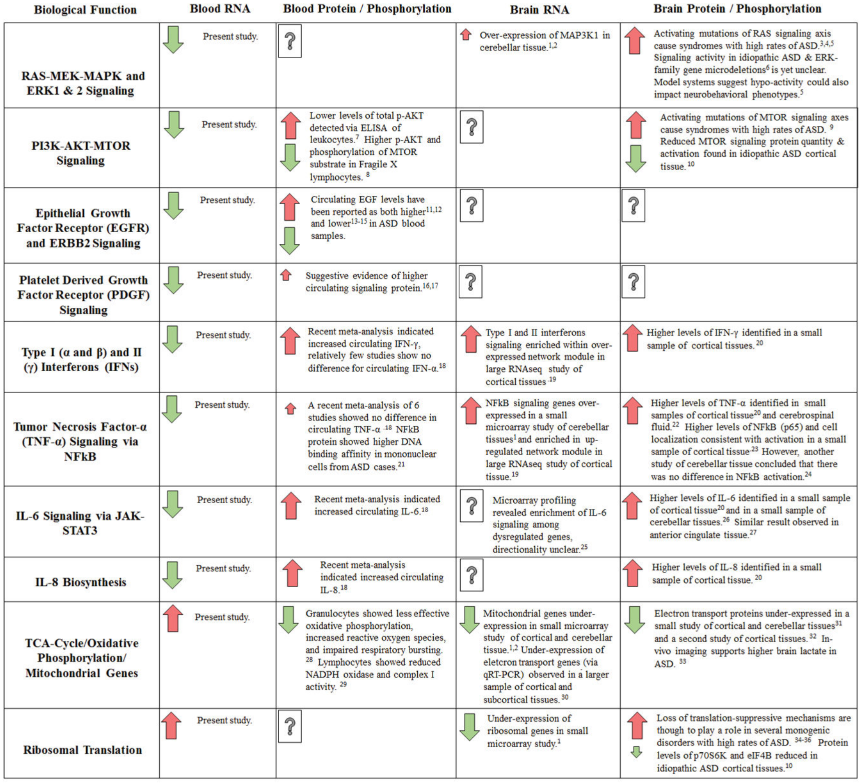
5 GWAS Hits: *ALDH1A1, C9orf72, LYRM1, MLKL, TFEC*

2 SFARIgene Candidates: *DAPP1, DOCK4*

Figure 2.

ASD-Associated Gene Co-expression Network Modules.

Gene co-expression network analysis performed on non-SVA-corrected data identified thirty network modules when ASD cases and comparison subjects were analyzed together. ASD-associated modules were identified using linear mixed models (as described in the single-gene analysis) to predict module eigengenes; five modules showed a significant association with ASD (Benjamini-Hochberg-corrected $p < 0.05$). Modules with higher eigengene values among ASD cases are described as over-expressed (left side) and those with lower eigengene values among ASD cases are described as under-expressed (right side). Within each colored panel, we indicate the module color name, the threshold of its significance for association with ASD, the number of positively and negatively loading genes, and functional and cell-type enrichments based on hypergeometric test statistics (Benjamini-Hochberg $p <$ 0.05; Full Results in Supplementary Table 9). We also include the names of genes that showed ASD GWAS signal at an uncorrected gene-level $p < 0.05$ in a recent meta-analysis. The symbol * beside the number of GWAS signal genes denotes whether the network module showed significant GWAS signal enrichment based on a quantitative permutation testing (see Methods). We also indicate the identities of SFARIgene candidate-genes contained within each module. Beside each colored panel is a plot depicting the 25 most highly inter-correlated genes for that network module; the 5 most highly correlated "hub" genes are depicted in the center. Each gene is depicted as a small colored circle, with the top right quarter indicating the relative over- or under-expression of that gene in ASD cases based on the single-gene covariate-controlled mega-analysis analysis; dark red indicates highly significant over-expression (FDR $q < 0.05$), while light red indicates nominal over-expression (uncorrected $p < 0.05$). The same relationships can be understood for dark green and light green, with respect to under-expressed genes in ASD. For each gene, the bottom left quarter of its circle was colored pink to indicate a gene showing nominal GWAS signal.



Biological Function	Blood RNA	Blood Protein / Phosphorylation	Brain RNA	Brain Protein / Phosphorylation
RAS-MEK-MAPK and ERK1 & 2 Signaling	↓ Present study.	?	↑ Over-expression of MAP3K1 in cerebellar tissue.[1,2]	↑ Activating mutations of RAS signaling axis cause syndromes with high rates of ASD.[3,4,5] Signaling activity in idiopathic ASD & ERK-family gene microdeletions[6] is yet unclear. Model systems suggest hypo-activity could also impact neurobehavioral phenotypes.[5]
PI3K-AKT-MTOR Signaling	↓ Present study.	↑ ↓ Lower levels of total p-AKT detected via ELISA of leukocytes.[7] Higher p-AKT and phosphorylation of MTOR substrate in Fragile X lymphocytes. [8]	?	↑ ↓ Activating mutations of MTOR signaling axes cause syndromes with high rates of ASD. [9] Reduced MTOR signaling protein quantity & activation found in idiopathic ASD cortical tissue. [10]
Epithelial Growth Factor Receptor (EGFR) and ERBB2 Signaling	↓ Present study.	↑ ↓ Circulating EGF levels have been reported as both higher[11,12] and lower[13-15] in ASD blood samples.	?	?
Platelet Derived Growth Factor Receptor (PDGF) Signaling	↓ Present study.	↑ Suggestive evidence of higher circulating signaling protein.[16,17]	?	?
Type I (α and β) and II (γ) Interferons (IFNs)	↓ Present study.	↑ Recent meta-analysis indicated increased circulating IFN-γ, relatively few studies show no difference for circulating IFN-α.[18]	↑ Type I and II interferons signaling enriched within over-expressed network module in large RNAseq study of cortical tissues .[19]	↑ Higher levels of IFN-γ identified in a small sample of cortical tissues. [20]
Tumor Necrosis Factor-α (TNF-α) Signaling via NFkB	↓ Present study.	↑ A recent meta-analysis of 6 studies showed no difference in circulating TNF-α .[18] NFkB protein showed higher DNA binding affinity in mononuclear cells from ASD cases.[21]	↑ NFkB signaling genes over-expressed in a small microarray study of cerebellar tissues[1] and enriched in up-regulated network module in large RNAseq study of cortical tissue.[19]	↑ Higher levels of TNF-α identified in small samples of cortical tissue[20] and cerebrospinal fluid.[22] Higher levels of NFkB (p65) and cell localization consistent with activation in a small sample of cortical tissue.[23] However, another study of cerebellar tissue concluded that there was no difference in NFkB activation.[24]
IL-6 Signaling via JAK-STAT3	↓ Present study.	↑ Recent meta-analysis indicated increased circulating IL-6.[18]	? Microarray profiling revealed enrichment of IL-6 signaling among dysregulated genes, directionality unclear.[25]	↑ Higher levels of IL-6 identified in a small sample of cortical tissue[20] and in a small sample of cerebellar tissues.[26] Similar result observed in anterior cingulate tissue.[27]
IL-8 Biosynthesis	↓ Present study.	↑ Recent meta-analysis indicated increased circulating IL-8.[18]	?	↑ Higher levels of IL-8 identified in a small sample of cortical tissue. [20]
TCA-Cycle/Oxidative Phosphorylation/ Mitochondrial Genes	↑ Present study.	↓ Granulocytes showed less effective oxidative phosphorylation, increased reactive oxygen species, and impaired respiratory bursting. [28] Lymphocytes showed reduced NADPH oxidase and complex I activity. [29]	↓ Mitochondrial genes under-expression in small microarray study of cortical and cerebellar tissue.[1,2] Under-expression of eletcron transport genes (via qRT-PCR) observed in a larger sample of cortical and subcortical tissues. [30]	↓ Electron transport proteins under-expressed in a small study of cortical and cerebellar tissues[31] and a second study of cortical tissues. [32] In-vivo imaging supports higher brain lactate in ASD. [33]
Ribosomal Translation	↑ Present study.	?	↓ Under-expression of ribosomal genes in small microarray study.[1]	↑ ↓ Loss of translation-suppressive mechanisms are though to play a role in several monogenic disorders with high rates of ASD. [34-36] Protein levels of p70S6K and eIF4B reduced in idiopathic ASD cortical tissues.[10]

Figure 3.

Summary of The Present Study's Findings in Comparison with Previous Studies of Human Blood and Brain Tissues with Respect to ASDs.

We conducted a literature search and attempted to compare our blood transcriptomic findings for various biological functions against previous studies that reported on the expression of either RNA or protein markers relevant to those functions in human blood or brain tissue, comparing ASD samples (including genetic syndromes with high rates of ASD) with unaffected comparison samples. All enumerated references supporting this table can be found in Supplementary Table 12. With respect to protein markers, we considered studies that examined the measured levels of circulating protein (e.g., blood cytokine or growth factor studies) or the relative quantity of a protein in cell or tissue, or the relative activation of protein signaling (e.g., quantity or proportion of phosphorylated protein). When clear directional findings were apparent from the reviewed literature, we denoted the conclusions with up or down arrows (reflecting ASD samples relative to controls) and provide brief descriptions of what was show, with supporting citations. When the literature clearly supports the identification of both increased and decreased activity of a biological function in ASD, we denoted this by including both up and down arrows, and attempt to provide more information on factors that might account for these findings. When insufficient evidence was found to draw a conclusion, we denoted this with a question mark. All supporting citations are discussed more thoroughly in the Supplementary Materials.

Table 1

Blood-Based Microarray Studies of Autism Spectrum Disorder Included in Mega-Analysis.

Study	Array Type	Cases (*n*)	Controls (*n*)	% Female	Age in Years (mean ± *s.d.)*	Sample Type	Predominant Ancestry, %	Genes Analyzed	Additional Sample Information†
Alter et al., 2011 GSE25507	Affymetrix U133 Plus 2.0	82	64	0%	6.6 ± 2.4	Lymphocytes	European, 100%	20 767	Phoenix area, Arizona, United States. Excluded CNS abnormalities, excluded known genetic or metabolic disorders, excluded subjects with Asperger's Syndrome, Autism Spectrum Disorder phenotypes, and signs of developmental regression. Genetic screening for chromosomal band abnormalities and Fragile X disorder. Samples drawn in Spring/Summer months.
‡ CHARGE Study Enstrom et al., 2009 Hertz-Picciotto et al., 2006 Stamova et al., 2011 Tian et al., 2011	Affymetrix U133 Plus 2.0	118	90	14.4%	3.7 ± 0.8	Whole Blood	European (only), 49.0% European identifying also as Latino, 81.0%	20 767	Davis area, California, United States. Included subjects wee 2–5 years of age. Cases met DSM-IV defined Autism based on administration of ADOS and ADI-R. Controls were age-, sex-matched and drawn from random sample of all births in 22-county catchment area of cases. No exclusions based on medical or genetic syndromes.
Glatt et al., 2012	Illumina WG-6 v3	173	159	30.7%	1.9 ± 0.8	Whole Blood	European, 100%	20,624	San Diego area, California, United States. Designed as a population-based study. Included subjects with Autism, Autism Spectrum Disorder, Asperger's Disorder, and Pervasive Developmental Disorder – Not Otherwise Specified. Did not exclude syndromic forms of ASD, cases featuring developmental regression, or cases based on medical comorbidity
Gregg et al., 2008 GSE6575	Affymetrix U133 Plus 2.0	35	11	17.3%	NA	Whole Blood	European, NA	20 767	Davis area, California, United States. Included subjects with Autism and Autism Spectrum Disorder. About half of cases showed signs of developmental regression.
Kong et al., 2012 GSE18123	Affymetrix U133 Plus 2.0	61	38	0%	8.3 ± 3.9	Whole Blood	European, 83.3%	20 767	Boston area, Massachusetts, United States. Included subjects with Autism, Autism Spectrum Disorder, Asperger's Disorder, and Pervasive Developmental Disorder – Not Otherwise Specified. Included subjects with comorbidities

Study	Array Type	Cases (*n*)	Controls (*n*)	% Female	Age in Years (mean ± *s.d.)*	Sample Type	Predominant Ancestry, %	Genes Analyzed	Additional Sample Information†
									including psychiatric and neurological disorders, auto-immune disorders, and gastrointestinal disorders.
Kong et al., 2012 GSE18123	Affymetrix HG 1.0 ST	104	68	30.8%	7.9 ± 4.3	Whole Blood	European 78.0%	24,515	Same as Above.
Kong et al., 2013	Affymetrix HG 1.0 ST	53	17	21.4%	9.6 ± 3.7	Whole Blood	European, 81.0%	24,515	Autism blood samples were obtained from the Simon's Simplex Collection. No known cases of genetic syndromes were included among these samples. Excluded subjects with chronic disease, such as infectious disease, diabetes, cardiovascular disease, and developmental disorder or neurological disorder. Control samples obtain from well-child visits in the Boston Massachusetts area, United States.
Total: 7	--	626	447	19.3%	5.1 ± 3.8	--	European, 92.3%	21 968 [*]	

Excluded studies with exclusion rationale are identified in Supplementary Table 2.

NA indicates data that were not specifically available from public databases or corresponding authors; these were coded as missing data for the purposes of analyses.

[†] We attempted to include available information pertaining to the study subjects, with particular emphasis on ASD phenotypic features and study exclusion criteria. All details were obtained from the referenced articles. For instances where specific details are not reported, readers can infer that this information was not included in the published studies.

[‡] Subsets of the CHARGE Study sample have been described in previous reports. The present study used the full CHARGE sample for which whole blood microarray data were available, which was larger than the subsets described previously.

[*] 21,969 reflects the total number of HGNC symbols that were represented in at least two of the studies identified above in which statistical models converged to produce a valid result for the mega-analysis of the full sample. Failures to converge were related to the number of missing values for a particular gene's expression level. Relatively fewer models converged to produce valid results in the sex-specific analyses because the number of genes with a higher proportion of missing data was higher in comparison to the total number of samples for analysis.

References

Alter, MD, Kharkar, R, Ramsey, KE, Craig, DW, Melmed, RD, Grebe, TA, Bay, RC, Ober-Reynolds, S, Kirwan, J, Jones, JJ, Turner, JB, Hen, R, Stephan, DA. 2011. Autism and increased paternal age related changes in global levels of gene expression regulation. PLoS One 6: e16715.

Enstrom, AM, Lit, L, Onore, CE, Gregg, JP, Hansen, RL, Pessah, IN, Hertz-Picciotto, I, Van de Water, JA, Sharp, FR, Ashwood, P. 2009. Altered gene expression and function of peripheral blood natural killer cells in children with autism. Brain. Behav. Immun. 23: 124–33.

Glatt, SJ, Tsuang, MT, Winn, M, Chandler, SD, Collins, M, Lopez, L, Weinfeld, M, Carter, C, Schork, N, Pierce, K, Courchesne, E. 2012. Blood-based gene expression signatures of infants and toddlers with autism. J. Am. Acad. Child Adolesc. Psychiatry 51.

Gregg, JP, Lit, L, Baron, CA, Hertz-Picciotto, I, Walker, W, Davis, RA, Croen, LA, Ozonoff, S, Hansen, R, Pessah, IN, Sharp, FR. 2008. Gene expression changes in children with autism. Genomics 91: 22–9.

Hertz-Picciotto, I, Croen, LA, Hansen, R, Jones, CR, van de Water, J, Pessah, IN. 2006. The CHARGE study: an epidemiologic investigation of genetic and environmental factors contributing to autism. Environ. Health Perspect. 114: 1119–25.

Kong, SW, Collins, CD, Shimizu-Motohashi, Y, Holm, IA, Campbell, MG, Lee, I-H, 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.

Kong, SW, Shimizu-Motohashi, Y, Campbell, MG, Lee, IH, Collins, CD, Brewster, SJ, Holm, I a, Rappaport, L, Kohane, IS, Kunkel, LM. 2013. Peripheral blood gene expression signature differentiates children with autism from unaffected siblings. Neurogenetics 14: 143–52.

Stamova, B, Green, PG, Tian, Y, Hertz-Picciotto, I, Pessah, IN, Hansen, R, Yang, X, Teng, J, Gregg, JP, Ashwood, P, Van de Water, J, Sharp, FR. 2011. Correlations between gene expression and mercury levels in blood of boys with and without autism. Neurotox. Res. 19: 31–48.

Tian, Y, Green, PG, Stamova, B, Hertz-Picciotto, I, Pessah, IN, Hansen, R, Yang, X, Gregg, JP, Ashwood, P, Jickling, G, Van de Water, J, Sharp, FR. 2011. Correlations of gene expression with blood lead levels in children with autism compared to typically developing controls. Neurotox. Res. 19: 1–13.

Table 2

Results of Machine Learning Classification Analysis.

Machine Type	Training Set (Boot67) Cross-Validation Accuracy	Validation Set Area Under ROC Curve	Validation Set Sensitivity	Validation Set Specificity
Linear Kernel SVM	0.72 ± 0.02	0.69 ± 0.03	0.65 ± 0.05	0.66 ± 0.05
Random Forests	0.71 ± 0.01	0.67 ± 0.03	0.63 ± 0.04	0.65 ± 0.04
Artificial Neural Networks	0.71 ± 0.02	0.69 ± 0.02	0.65 ± 0.05	0.66 ± 0.04