

## Original Investigation

# Prediction of Autism by Translation and Immune/Inflammation Coexpressed Genes in Toddlers From Pediatric Community Practices

Tiziano Pramparo, PhD; Karen Pierce, PhD; Michael V. Lombardo, PhD; Cynthia Carter Barnes, PhD; Steven Marinero, BS; Clelia Ahrens-Barbeau, BS; Sarah S. Murray, PhD; Linda Lopez, MA; Ronghui Xu, PhD; Eric Courchesne, PhD

 Supplemental content at [jamapsychiatry.com](http://jamapsychiatry.com)

**IMPORTANCE** The identification of genomic signatures that aid early identification of individuals at risk for autism spectrum disorder (ASD) in the toddler period remains a major challenge because of the genetic and phenotypic heterogeneity of the disorder. Generally, ASD is not diagnosed before the fourth to fifth birthday.

**OBJECTIVE** To apply a functional genomic approach to identify a biologically relevant signature with promising performance in the diagnostic classification of infants and toddlers with ASD.

**DESIGN, SETTING, AND PARTICIPANTS** Proof-of-principle study of leukocyte RNA expression levels from 2 independent cohorts of children aged 1 to 4 years (142 discovery participants and 73 replication participants) using Illumina microarrays. Coexpression analysis of differentially expressed genes between Discovery ASD and control toddlers were used to define gene modules and eigengenes used in a diagnostic classification analysis. Independent validation of the classifier performance was tested on the replication cohort. Pathway enrichment and protein-protein interaction analyses were used to confirm biological relevance of the functional networks in the classifier. Participant recruitment occurred in general pediatric clinics and community settings. Male infants and toddlers (age range, 1-4 years) were enrolled in the study. Recruitment criteria followed the 1-Year Well-Baby Check-Up Approach. Diagnostic judgment followed *DSM-IV-TR* and Autism Diagnostic Observation Schedule criteria for autism. Participants with ASD were compared with control groups composed of typically developing toddlers as well as toddlers with global developmental or language delay.

**MAIN OUTCOMES AND MEASURES** Logistic regression and receiver operating characteristic curve analysis were used in a classification test to establish the accuracy, specificity, and sensitivity of the module-based classifier.

**RESULTS** Our signature of differentially coexpressed genes was enriched in translation and immune/inflammation functions and produced 83% accuracy. In an independent test with approximately half of the sample and a different microarray, the diagnostic classification of ASD vs control samples was 75% accurate. Consistent with its ASD specificity, our signature did not distinguish toddlers with global developmental or language delay from typically developing toddlers (62% accuracy).

**CONCLUSIONS AND RELEVANCE** This proof-of-principle study demonstrated that genomic biomarkers with very good sensitivity and specificity for boys with ASD in general pediatric settings can be identified. It also showed that a blood-based clinical test for at-risk male infants and toddlers could be refined and routinely implemented in pediatric diagnostic settings.

*JAMA Psychiatry*. 2015;72(4):386-394. doi:10.1001/jamapsychiatry.2014.3008  
Published online March 4, 2015.

**Author Affiliations:** Author affiliations are listed at the end of this article.

**Corresponding Author:** Eric Courchesne, PhD, UC San Diego Autism Center of Excellence, Department of Neuroscience, University of California, San Diego School of Medicine, La Jolla, CA 92093 ([ecourchesne@ucsd.edu](mailto:ecourchesne@ucsd.edu)).

**A**utism spectrum disorder (ASD) is a neurodevelopmental disorder of complex etiology with early onset and generally is not diagnosed before a median age of 53 months in the United States.<sup>1</sup> Early and long-term intervention is the most effective strategy to reduce or reverse the core features of toddlers and children with autism.<sup>2,3</sup> While current treatment strategies are rapidly progressing for a more effective intervention in autism and other developmental disorders,<sup>4</sup> the hunt for biological markers or genetic signatures is ongoing. Together with the complex nature of the disorder, the identification of biomarkers or molecular signatures is limited by the inaccessibility of neural tissue and the few young postmortem samples available.

Peripheral blood of living individuals is a preferable and more accessible tissue for such screening. It is expected to carry autism-relevant signatures that can be used to detect the disorder at very young ages and might also reflect aspects of the disrupted biology underlying neural defects.

A few studies<sup>5-7</sup> have investigated cohorts of blood-derived samples, both *in vivo* and *in vitro*, to describe sets of differentially expressed (DE) genes that distinguish individuals with ASD vs control subjects. The largest *in vivo* study<sup>6</sup> achieved a validated classification accuracy of 68% in children with a mean age of approximately 8 years. Despite these promising efforts, additional studies are needed to support or improve genetic signatures with high specificity and sensitivity and, most important, to push the prediction power at very young ages, when intervention is most effective.<sup>2,3</sup> In the long run, a practical clinical test will require these signatures to be effective in the general pediatric population and not just in preselected syndromic patients or patients with ASD from multiplex families.

Using a systems biology approach, we conducted a proof-of-principle study using leukocyte gene expression aimed to identify a genomic signature that classified with good accuracy 2 independent cohorts of infants and toddlers with ASD (mean age, approximately 2 years) recruited through community pediatric clinics and other community sources. Several genomic signatures are expected to coexist. However, until very large sample sizes of individuals of different genetic backgrounds at young ages become available, our signature represents an unprecedented study outcome that is based on a general pediatric population. With the identification of consistent dysregulated gene pathways and gene sets with predictive roles in ASD, it is expected that biomarker discovery with high specificity is possible and that a blood-based clinical test can be implemented in a routine diagnostic setting.

## Methods

### Participant Recruitment, Tracking, and Developmental Evaluation

Given the substantial 4:1 male to female bias in ASD and several reasons to suspect that potentially important sex differential factors underlie etiological aspects of autism (eg, in the review by Schaafsma and Pfaff<sup>8</sup> on this topic), we chose to focus on boys only to reduce the potential increase in

genomic heterogeneity that would accompany a mixed-sex design. Included in the study were 220 participants aged 1 to 4 years, including 147 toddlers in a discovery sample (91 ASD and 56 control) and 73 toddlers (44 ASD and 29 control) in a replication sample. Sample collection occurred from 2009 to 2011, and diagnostic evaluation occurred from 2009 to 2013. The replication sample largely overlapped with individuals in our previous pilot study<sup>5</sup> of leukocyte gene expression, but the discovery toddlers represent a new and independent sample.

All toddlers were developmentally evaluated by a PhD-level psychologist (C.C.B.), and those younger than 3 years at the time of blood draw were tracked every 6 months until their third birthday, when a final diagnosis was given. Only toddlers with a provisional or confirmed ASD diagnosis were included in the present study. Toddlers were recruited via the 1-Year Well-Baby Check-Up Approach from community pediatric clinics.<sup>2</sup> This approach enables a general naturalistic population screening approach for prospective study of ASD, typically developing patients, and patients with contrast developmental delay (eg, language, global developmental, or motor delay). In this approach, parents of toddlers completed a broadband developmental screen at their pediatrician's office, and toddlers were referred, evaluated, and tracked over time. This approach provided an unbiased recruitment of toddlers representing a wide range and variety of ability and disability. Blood samples for gene expression and DNA analysis were collected from a subset of participants at the time of referral, regardless of referral reason, and before final diagnostic evaluations. No blood draws were performed if participants showed signs of influenza, a cold, or infections or if any illnesses were present or suspected 72 hours before visits. Every participant was evaluated using multiple tests, including the appropriate module of the Autism Diagnostic Observation Schedule<sup>9,10</sup> and the Mullen Scales of Early Learning.<sup>11</sup> Diagnoses were determined via these assessments and the *DSM-IV-TR*.<sup>12</sup> Parents were interviewed using the Vineland Adaptive Behavior Scales<sup>13</sup> and underwent a medical history interview. Both the discovery and replication cohorts included individuals with ASD and control participants. The control group was composed of typically developing toddlers and toddlers with contrast developmental delay (Table). Institutional review board approval from the University of California, San Diego, was obtained for the study. Written informed consent was obtained from the parents of the participants. Additional methodological information is provided in the eMethods in Supplement 1.

### Blood Sample Collection, RNA Extraction, and Processing

Four to six milliliters of blood was collected into ethylenediaminetetraacetic-coated tubes from all toddlers, passed over a filter (LeukoLOCK; Ambion) to capture and stabilize leukocytes, and immediately placed in a  $-20^{\circ}\text{C}$  freezer. Total RNA was extracted following standard procedures and manufacturer's instructions (Ambion). The RNA samples in the discovery data set were tested using one platform (HT-12; Illumina), while the RNA samples in the replication data set were tested using another platform (WG-6; Illumina). Several quality cri-

Table. Summary of Participant Characteristics and Clinical Information

Variable	Mean (SD)					
	Discovery			Replication		
	ASD (n = 87)	TD (n = 41)	Contrast DD <sup>a</sup> (n = 14)	ASD (n = 44)	TD (n = 25)	Contrast DD <sup>a</sup> (n = 4)
Age, y	2.3 (0.7)	2.0 (0.9)	1.5 (0.6)	2.3 (0.8)	1.6 (0.7)	1.2 (0.2)
Mullen Scales of Early Learning <i>t</i> score						
Visual reception	39.7 (11.0)	59.0 (10.3)	48.1 (9.0)	40.6 (13.6)	51.6 (10.2)	44.3 (4.5)
Fine motor	37.3 (12.2)	55.9 (9.1)	55.8 (8.4)	40.1 (16.0)	57.5 (8.5)	55.7 (2.9)
Receptive language	29.1 (12.0)	52.4 (8.3)	46.9 (8.5)	31.6 (16.1)	50.7 (10.2)	36.7 (4.9)
Expressive language	29.1 (11.4)	53.7 (9.5)	46.3 (7.9)	31.4 (16.4)	52.0 (8.6)	41.0 (2.6)
Early learning composite	71.0 (16.2)	110.5 (12.4)	98.7 (11.4)	76.1 (21.6)	106.0 (12.9)	89.3 (6.7)
Autism Diagnostic Observation Schedule score <sup>b</sup>						
Communication social/social affect	15.0 (3.9)	2.1 (1.7)	4.2 (3.7)	12.8 (4.8)	2.4 (2.2)	5.0 (5.0)
Restricted, repetitive behaviors	4.1 (1.9)	0.3 (0.5)	0.6 (1.1)	2.5 (1.6)	0.3 (0.4)	0.7 (1.2)
<b>Total</b>	<b>19.1 (4.7)</b>	<b>2.4 (1.9)</b>	<b>4.8 (4.6)</b>	<b>15.3 (5.4)</b>	<b>2.7 (2.3)</b>	<b>5.7 (6.0)</b>
Vineland Adaptive Behavior Scales adaptive behavior composite score <sup>c</sup>	82.2 (9.4)	101.6 (9.3)	92.4 (7.6)	83.6 (14.1)	100.8 (7.3)	95.0 (1.0)

Abbreviations: ASD, autism spectrum disorder; DD, developmental delay; TD, typically developing.

<sup>a</sup> Toddlers in the Contrast DD category included language delay (9 discovery and 2 replication), global developmental delay (1 replication), radiological abnormality (1 discovery and 1 replication), premature birth but testing within the normal range on standardized tests (2 discovery), socially emotionally delayed (1 discovery), and prenatal drug exposure (1 discovery).

<sup>b</sup> All toddlers received either the toddler module or module 1 or 2 depending on age and verbal ability at the time of testing.

<sup>c</sup> In the discovery sample, 64% of individuals with ASD had Autism Diagnostic Observation Schedule (ADOS) toddler module, 31% had ADOS module 1, and 5% had ADOS module 2. In the replication sample, 32% of individuals with ASD had ADOS toddler module, 48% had ADOS module 1, and 20% had ADOS module 2.

teria were used to exclude low-quality arrays as previously described.<sup>14,15</sup> Five low-quality arrays in the discovery data set were identified and excluded from statistical analyses. All arrays from the replication data set were of high quality. Both the discovery and replication data sets underwent the same filtering and normalization steps. Final samples represented 87 toddlers with ASD and 55 control participants (total, 142 participants) in the discovery cohort and 44 toddlers with ASD and 29 control participants in the replication cohort (Table). Raw and normalized data are deposited in the Gene Expression Omnibus (GSE42133). Additional methodological information is provided in the eMethods in Supplement 1.

### Statistical and Bioinformatic Analysis

Figure 1 shows a schematic representation of the main statistical and bioinformatics analyses. Statistical analyses were performed on normalized and filtered expression data. Class comparison analysis was performed to identify DE genes using a standard univariate 2-sample *t* test model with 10 000 random permutations using a software package (BRB-Array Tools; Biometric Research Branch, National Cancer Institute). The significant threshold of univariate tests was .05 (Supplement 2). Differentially expressed genes were then used for enrichment pathway analysis using an available tool (Metacore; GeneGo) and coexpression analysis. A weighted gene network coexpression analysis package<sup>16,17</sup> was used to identify coexpression modules in an unsupervised fashion from DE genes (ie, clusters of DE genes that are tightly coexpressed across all discovery sample participants) and to calculate the first principal component of each module, herein called the module eigengene (ME). The ME is a value that

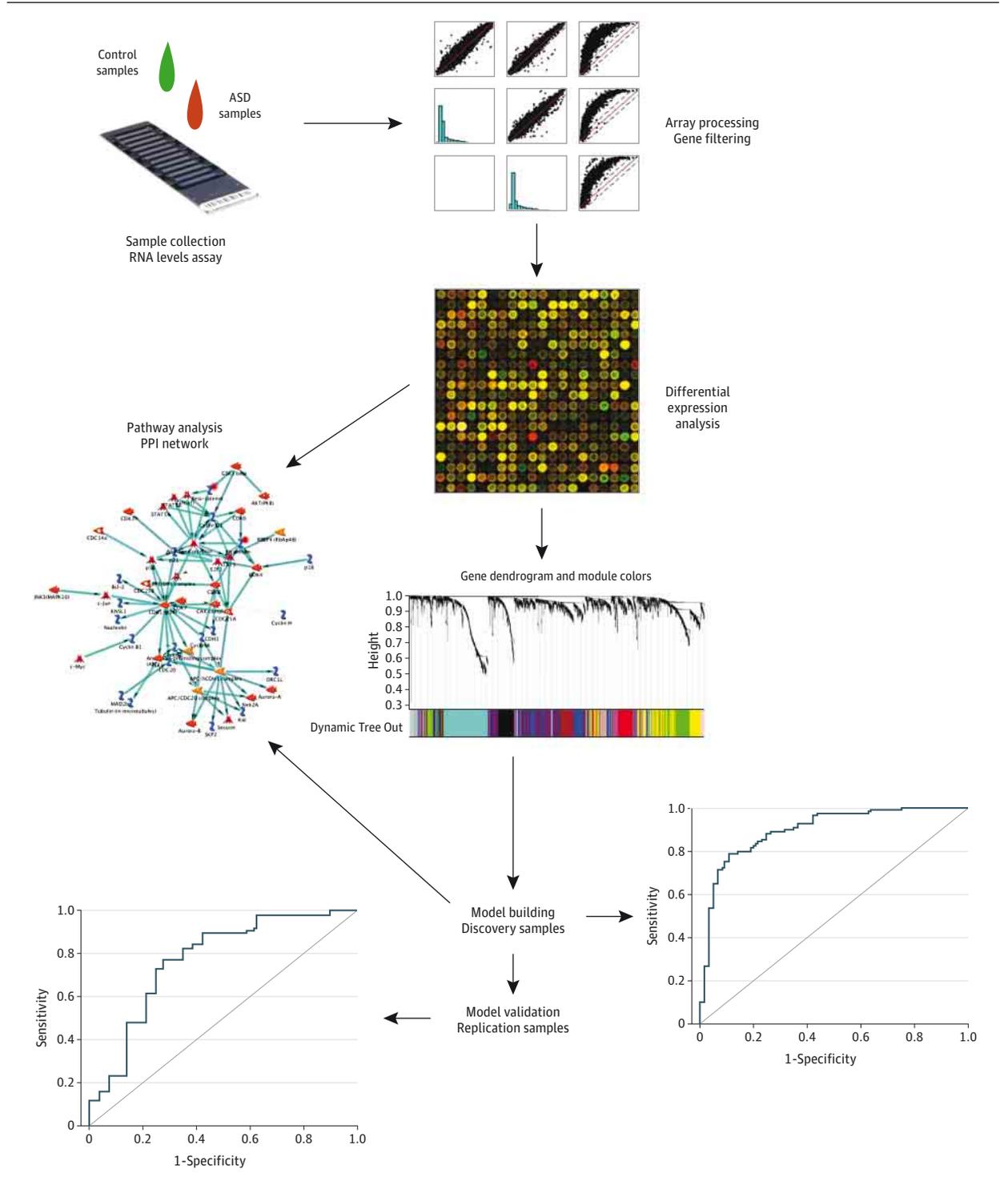
summarizes each module's expression profile and can be understood as a weighted average of the gene expression profiles within a module.<sup>18</sup>

Coexpression analysis was run by selecting the lowest power for which the scale-free topology fit index reached 0.90 (soft power threshold, 5.5) and by constructing a signed (ie, bi-directional) network with a hybrid dynamic branch-cutting method to assign individual genes to modules.<sup>19</sup> Hypergeometric probability was used to test the significance in gene overlap vs random gene sets of equal size. A software program (CNVision; Yale University) was used to call copy number variations (CNVs) in misclassified individuals with ASD as previously described.<sup>15,20</sup> Additional methodological information, including the differential expression analysis of cortical tissue, is provided in the eMethods in Supplement 1.

### Classification Analysis

Twelve MEs were obtained from the weighted gene network coexpression analysis of 2765 DE genes in the discovery sample. Feature selection<sup>6</sup> based on the MEs was achieved by running logistic regression on the ME values. We began by first identifying a pair of modules that performed best at distinguishing participants with ASD from control participants, followed by recursively adding one extra module at a time and retaining it if it increased performance. From this process, 4 of 12 modules were identified (M1 is blue, M2 is black, M3 is purple, and M4 is green yellow in eFigure 1 in Supplement 1) that displayed the best area under the curve performance. These 4 modules were further used to test the classification accuracy on the independent replication sample. To validate the accuracy of the classifier on an inde-

Figure 1. Diagram of the Study Design Shows the Main Statistical and Bioinformatics Analyses



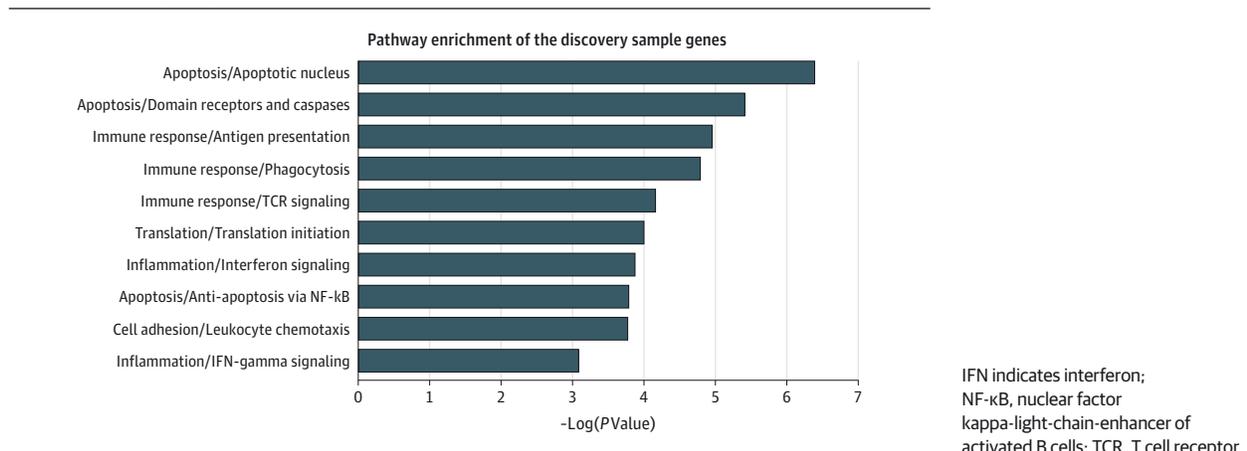
Microarray gene expression data from all available samples were processed to select for high-quality data and stronger gene expression signal. Processed data were first analyzed by differential expression and then by coexpression analysis to reduce dimensionality and leverage characteristics based on network biology. Feature selection and model building were performed on the discovery samples

only (training data). The model obtained from the training data was then applied, without any further modifications, to the replication samples (validation data set) to independently test classification performance. Network-based analyses were also applied to the classification signature to identify its functional characteristics. ASD indicates autism spectrum disorder; PPI, protein-protein interaction.

pendent set of samples (replication set), gene weights were computed from the discovery cohort as the correlation

between each gene in the 4 modules and their respective eigengene values. Weights were applied to the gene expres-

**Figure 2. Pathway Enrichment Using Metacore GeneGo of Differentially Expressed Genes From the Discovery Sample**



sion levels of each replication participant, and eigengenes were computed and used in logistic regression. A software package (caret; <http://caret.r-forge.r-project.org/>) and default settings were used to run the logistic regression function (glmnet) with repeated (3 times) 10-fold cross-validation on the training set only (discovery sample). The model obtained from the discovery data was then applied, without any further modifications, on the replication data to test classification performance. Because of the differences in microarray platforms, only 2070 of 2765 discovery DE genes (75%) were present in the replication data set, and only 678 of 762 four-module classifier genes (89%) were actually represented by replication MEs and used in the classification test of the replication participants. Clinical and magnetic resonance imaging characteristics between the correctly classified and misclassified groups (ASD and control) were compared in both the discovery and replication samples to determine if the classifier could be biased by differences in clinical and neuroanatomic characteristics. Results for the Mullen Scales of Early Learning, Autism Diagnostic Observation Schedule, and Vineland Adaptive Behavior Scales scores were compared. Residual brain volumes for total brain volume, cerebral white and gray matter, and cerebellar white and gray matter were also compared.

## Results

Most discovery and replication cohort members were of white race/ethnicity. Pearson  $\chi^2$  test showed no significant difference in racial/ethnic characteristics between individuals with ASD and control participants ( $\chi^2_5 = 7.98, P = .16$  for the discovery cohort and  $\chi^2_5 = 7.19, P = .21$  for the replication cohort). After filtering across all discovery cohort members, 12 088 gene probes were used for downstream analyses. Multivariable regression analysis showed no variance explained by differences in racial/ethnic characteristics between individuals with ASD and control participants, and 4% of variance was explained by age ( $P < .05$ ).

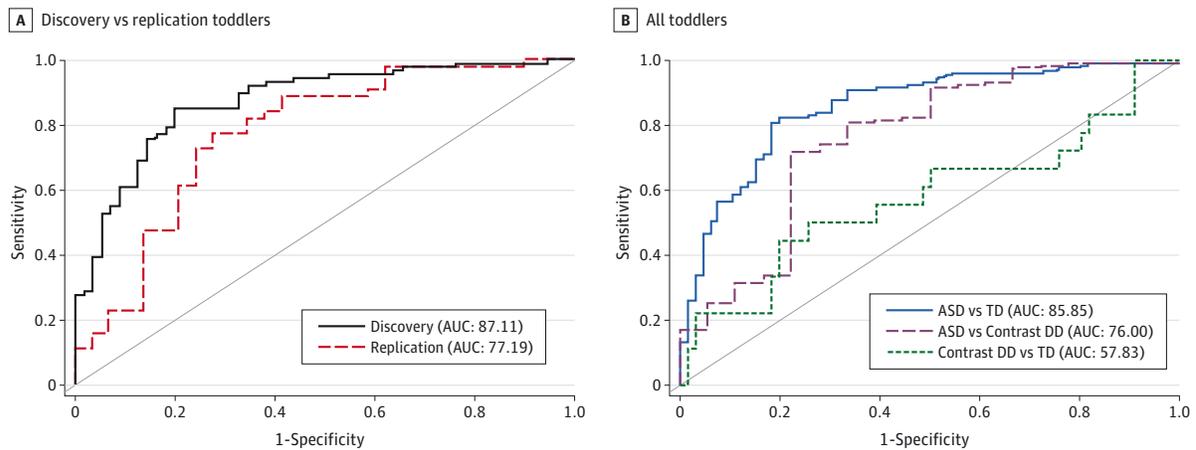
## Module-Based Classification Efficiently Distinguishes Individuals With ASD From Control Participants

Class comparison of discovery cohort members identified 2765 unique DE genes, with top enrichment in apoptosis, immune/inflammation response, and translation networks (Figure 2 and Supplement 2). Coexpression analysis identified 12 modules (eFigure 1 in Supplement 1), and eigengenes were calculated for each discovery cohort member and each module. Four modules' eigengenes were used in the classification analysis together with each individual's age as a predictor. These modules contained 762 unique genes (423 M1 genes, 191 M2 genes, 90 M3 genes, and 58 M4 genes) (Supplement 2). Logistic regression of diagnosis with age as the predictor produced an odds ratio of 1.07 (95% CI, 1.03-1.12;  $P < .05$ ), and classification without age was 3% to 4% less accurate. Logistic regression and receiver operating characteristic analyses displayed a high area under the curve in both the discovery cohort (training set on Illumina HT-12) and the replication cohort (independent test set on Illumina WG-6), with 83% and 75% classification accuracy, respectively (Figure 3 and Supplement 2). While the specificity remained high across the different class comparisons, the accuracy and sensitivity decreased as the sample size was reduced (Figure 3). We questioned whether misclassified individuals carried known pathogenic CNVs for failing correct diagnostic prediction. At a 0.5 threshold, 12 of 14 misclassified individuals with ASD were genotyped for CNV analyses. A rare CNV of known ASD etiology, CNTNAP2 duplication, was found in only one individual (eTable in Supplement 1). No clinical, behavioral, or magnetic resonance imaging-based measures indicated subphenotypic differences between correctly classified and misclassified individuals.

## Characterization of the Classification Signature

Metacore GeneGo analysis of the 4-module classifier displayed significant enrichment in translation and immune/inflammation genes (Figure 4 and Supplement 2). We sought to independently validate these findings by querying the DAPPLE database (<http://www.broadinstitute.org/mpg/dapple>), which looks for significant physical connectivity

Figure 3. Receiver Operating Characteristic (ROC) Curves and Extracted Values

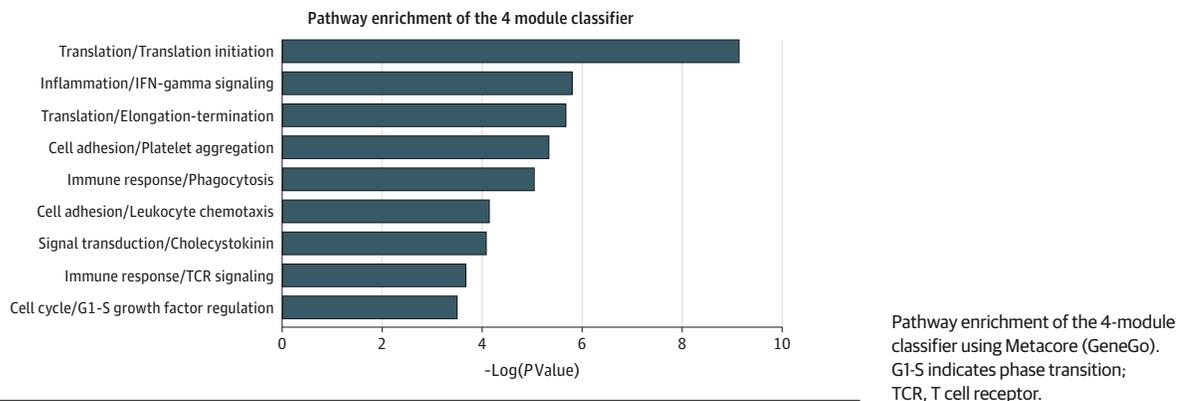


	Discovery	Replication	ASD vs TD	ASD vs Contrast DD	Contrast DD vs TD
Participants, No.	142	73	197	149	84
Best threshold	0.559	0.547	0.547	0.614	0.547
True-positive	74	34	108	94	8
True-negative	44	21	53	14	53
False-positive	11	8	13	4	13
False-negative	13	10	23	37	10
Accuracy, %	82.53	74.84	81.37	74.77	62.37
Specificity, %	80.00	72.41	80.30	77.77	80.30
Sensitivity, %	85.05	77.27	82.44	71.75	44.44

A, ROC curves and area under the curve (AUC) values from the classification of discovery (black) and replication (red) toddlers. B, ROC curves and AUC values from the classification of all toddlers in the different diagnostic subcategories. Blue indicates toddlers with autism spectrum disorder (ASD) vs typically

developing (TD) toddlers, thus excluding toddlers with contrast developmental delay (DD). Purple indicates toddlers with ASD vs toddlers with contrast DD. Green indicates toddlers with contrast DD vs TD toddlers. C, Confusion matrix and classification scores using the best threshold for each test.

Figure 4. Pathway Enrichment of the 4-Module Classifier



among proteins encoded by genes in loci associated with the disorder according to protein-protein interactions (PPIs) reported in the literature.<sup>21</sup> Indeed, these gene modules revealed a statistical enrichment for PPI ( $P < .001$ ). Using this PPI information, we created a classification network to map the genes with the highest number of PPIs. Consistent with enrichment findings of the 762-gene signature, the PPI gene list displayed translation initiation as the top process network ( $P = 4e-18$ ). A substantial number of ribosomal and transla-

tion genes were positioned at the center of the PPI network (Figure 5), suggesting that these central genes may have important regulative roles.

To address whether the PPI network also included genes that are potentially relevant to the disrupted biology of the ASD brain, we next performed differential expression analysis (eMethods in Supplement 1) using data from a recent post-mortem tissue study<sup>22</sup> and mapped the cortex-specific DE genes in the PPI network. We found a statistically significant



erature, especially compared with the performance of other tests applied to the young general pediatric population (as opposed to preselected syndromic patients or patients with ASD from multiplex families). For example, the Modified Checklist for Autism in Toddlers, a commonly used parent-report screen, has very low specificity (27%)<sup>23</sup> and positive predictive value (11%-54%) when used in general populations.<sup>24,25</sup> While important strides have been made to understand possible genetic risk factors in autism,<sup>26</sup> current DNA tests detect only rare autism cases and lack specificity<sup>27</sup> or confirm autism at older ages, and these tests have not been demonstrated to be effective in infants and toddlers with ASD.<sup>6</sup> Although the present study was designed and performed as proof of principle to stimulate research in the identification of an early biomarker of ASD from a general pediatric population, the candidate functional genomic signature reported herein has shown far greater diagnostic potential than other blood-based or behavior-based candidate classifiers in infants and toddlers with ASD. Nonetheless, the candidate genomic signature we describe represents only the first step toward a practical and accurate first-tier screen. Larger validation studies are needed, as are further studies of the specificity and sensitivity relative to other neurodevelopmental disorders.

The gene list of this new signature has little overlap with the candidate gene list in our group's previous, small pilot study.<sup>5</sup> This may be due to differences in the overall study design (eg, sample size, sex), feature selection strategy (fold-change vs module-based coexpression), and classification algorithm (machine learning vs standard logistic regression), as well as because of the clinical heterogeneity of the individuals investigated. While the individuals in the replication cohort in the present study largely overlapped with those in the previous pilot study, the module-based signature was derived from an independent, newly collected sample set and was tested on a more recent microarray platform. It is likely that higher concordance of findings between the 2 studies would occur if results were compared at the pathway level rather than the single-gene level. For example, both studies found signals related to genes involved in the immune response pathway. Dysregulation of immune/inflammation mechanisms has been described in a large number of autism studies.<sup>28</sup> To our knowledge, the present study is the first to detect strong significant dysregulation of immune and inflammation gene networks at approximately the age at the first emergence of the clinical risk signs of ASD. Blood cell-derived gene expression studies<sup>6,28</sup> of older children and adults with ASD also report dysregulation of immune/inflammation genes. Although evi-

dence of immune involvement has been argued to be a later secondary abnormality in ASD, there is no experimental evidence to favor that idea over the possibility that ASD may involve both prenatal immune (eg, maternal immune activation) and genetic factors.<sup>29,30</sup>

In addition, our signature revealed a central role of protein synthesis in the diagnostic classification of autism. The finding of translation genes at the core in the classification network is a strong reminder of the mechanism underlying the most common single-gene mutation in ASD.<sup>31,32</sup> Although it was not the focus of this study to address the level at which protein synthesis is altered, it seems plausible to think that global regulation of translation may be affected. A model of such global dysregulation may well explain the heterogeneity of gene networks and pathways that are involved and disrupted in autism<sup>33</sup> and this model has been proposed for transcription genes in children with idiopathic autism.<sup>34</sup> In contrast, we have not detected a statistically significant signal for synaptic pathways, although *de novo* mutations of genes with roles in synaptic function and localization have been identified<sup>35</sup> and represent a point of convergence in ASD.<sup>36</sup> We argue that the lack of significant signal for synaptic pathways may be due to the use of blood tissue, which is more likely to reveal changes in genes expressed at the systemic level rather than genes with high expression levels that are specific to neuronal tissue. However, we detected a significant overlap between genes in the PPI classifier network and DE genes in brain tissue from ASD cases, and a substantial proportion of these overlapping genes displayed a high number of PPIs (see the network core in Figure 5). Therefore, these results indicate that among genes in our *in vivo* blood classifier are those that can have effects on brain development, with functional enrichment in translation, immune, and cell cycle processes. The importance of blood as an *in vivo* measurement should be underscored because it is of substantial practical importance to the potential advancement of clinical uses for the expression signatures we report herein.

## Conclusions

In conclusion, knowledge of these common pathways and changes in hub gene connectivity patterns will facilitate research into biological targets for biotherapeutic intervention. The findings will aid the development of accurate biomarkers for detecting risk for ASD among infants in the general pediatric population.

### ARTICLE INFORMATION

**Submitted for Publication:** June 23, 2014; final revision received October 27, 2014; accepted October 28, 2014.

**Published Online:** March 4, 2015.

doi:10.1001/jamapsychiatry.2014.3008.

**Author Affiliations:** UC San Diego Autism Center of Excellence, Department of Neuroscience, University of California, San Diego School of Medicine, La Jolla (Pramparo, Pierce, Carter Barnes, Marinero, Ahrens-Barbeau, Lopez, Courchesne);

Autism Research Centre, Department of Psychiatry, University of Cambridge, Cambridge, United Kingdom (Lombardo); Department of Psychology, University of Cyprus, Nicosia, Cyprus (Lombardo); Center for Applied Neuroscience, University of Cyprus, Nicosia, Cyprus (Lombardo); Scripps Genomic Medicine and Scripps Translational Sciences Institute, La Jolla, California (Murray); Department of Pathology, University of California, San Diego, La Jolla (Murray); Department of Family and Preventive Medicine, University of California,

San Diego, La Jolla (Xu); Department of Mathematics, University of California, San Diego, La Jolla (Xu).

**Author Contributions:** Drs Pramparo and Courchesne had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. *Study concept and design:* Pierce, Courchesne. *Acquisition, analysis, or interpretation of data:* Pramparo, Pierce, Lombardo, Marinero, Courchesne.

*Drafting of the manuscript:* Pramparo, Pierce, Lombardo, Xu, Courchesne.

*Critical revision of the manuscript for important intellectual content:* Pierce, Lombardo, Xu, Courchesne.

*Statistical analysis:* Xu.

*Administrative, technical, or material support:* Pierce, Carter Barnes, Ahrens-Barbeau, Murray, Lopez. *Study supervision:* Pierce, Xu, Courchesne.

**Conflict of Interest Disclosures:** Drs Pramparo and Courchesne reported a patent pending that includes data from this study. No other disclosures were reported.

**Funding/Support:** This research was supported by grants P50-MH081755 and R01-MH036840 (Dr Courchesne), R01-MH080134 (Dr Pierce), 1U54RR025204-01 (Mr Marinero), and 1UL1RR031980-01 (Dr Xu) from the National Institutes of Health and by grant KL2TRO0099 from the University of California, San Diego Clinical and Translational Research Institute (Dr Pramparo).

**Role of the Funder/Sponsor:** The funding sources had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

**Additional Contributions:** Roxana Hazin, PhD, at the UC San Diego Autism Center of Excellence helped with participant recruitment. We thank all the families for making this study possible.

## REFERENCES

- Developmental Disabilities Monitoring Network Surveillance Year 2010 Principal Investigators; Centers for Disease Control and Prevention (CDC). Prevalence of autism spectrum disorder among children aged 8 years. *MMWR Surveill Summ*. 2014; 63(2):1-21.
- Pierce K, Carter C, Weinfeld M, et al. Detecting, studying, and treating autism early. *J Pediatr*. 2011; 159(3):458-465.e1-6. doi:10.1016/j.jpeds.2011.02.036.
- Warren Z, McPheeters ML, Sathe N, Foss-Feig JH, Glasser A, Veenstra-Vanderweele J. A systematic review of early intensive intervention for autism spectrum disorders. *Pediatrics*. 2011; 127(5):e1303-e1311. doi:10.1542/peds.2011-0426.
- Wallace KS, Rogers SJ. Intervening in infancy: implications for autism spectrum disorders. *J Child Psychol Psychiatry*. 2010; 51(12):1300-1320.
- Glatt SJ, Tsuang MT, Winn M, et al. Blood-based gene expression signatures of infants and toddlers with autism. *J Am Acad Child Adolesc Psychiatry*. 2012; 51(9):934-944.e2. doi:10.1016/j.jaac.2012.07.007.
- Kong SW, Collins CD, Shimizu-Motohashi Y, et al. Characteristics and predictive value of blood transcriptome signature in males with autism spectrum disorders. *PLoS One*. 2012; 7(12):e49475. doi:10.1371/journal.pone.0049475.
- Hu VW, Lai Y. Developing a predictive gene classifier for autism spectrum disorders based upon differential gene expression profiles of phenotypic subgroups. *N Am J Med Sci (Boston)*. 2013; 6(3). doi:10.7156/najms.2013.0603107.
- Schaafsma SM, Pfaff DW. Etiologies underlying sex differences in autism spectrum disorders. *Front Neuroendocrinol*. 2014; 35(3):255-271.
- Luyster R, Gotham K, Guthrie W, et al. The Autism Diagnostic Observation Schedule-Toddler Module. *J Autism Dev Disord*. 2009; 39(9):1305-1320.
- Lord C, Risi S, Lambrecht L, et al. The Autism Diagnostic Observation Schedule-Generic. *J Autism Dev Disord*. 2000; 30(3):205-223.
- Mullen EM. *Mullen Scales of Early Learning*. Circle Pines, MN: American Guidance Service Inc; 1995.
- American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 4th ed, text revision. Washington, DC: American Psychiatric Association; 2000.
- Sparrow SS, Cicchetti DV, Balla DA. *Vineland Adaptive Behavior Scales*. 2nd ed. San Antonio, TX: Pearson Education Inc; 2005.
- Chow ML, Li HR, Winn ME, et al. Genome-wide expression assay comparison across frozen and fixed postmortem brain tissue samples. *BMC Genomics*. 2011; 12:449.
- Chow ML, Pramparo T, Winn ME, et al. Age-dependent brain gene expression and copy number anomalies in autism suggest distinct pathological processes at young versus mature ages. *PLoS Genet*. 2012; 8(3):e1002592. doi:10.1371/journal.pgen.
- Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008; 9:559.
- Langfelder P, Horvath S. Eigengene networks for studying the relationships between co-expression modules. *BMC Syst Biol*. 2007; 1:54.
- Cai C, Langfelder P, Fuller TF, et al. Is human blood a good surrogate for brain tissue in transcriptional studies? *BMC Genomics*. 2010; 11:589.
- Pramparo T, Libiger O, Jain S, et al. Global developmental gene expression and pathway analysis of normal brain development and mouse models of human neuronal migration defects. *PLoS Genet*. 2011; 7(3):e1001331. doi:10.1371/journal.pgen.1001331.
- Sanders SJ, Ercan-Sencicek AG, Hus V, et al. Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. *Neuron*. 2011; 70(5):863-885.
- Rossin EJ, Lage K, Raychaudhuri S, et al; International Inflammatory Bowel Disease Genetics Consortium. Proteins encoded in genomic regions associated with immune-mediated disease physically interact and suggest underlying biology. *PLoS Genet*. 2011; 7(1):e1001273. doi:10.1371/journal.pgen.1001273.
- Voineagu I, Wang X, Johnston P, et al. Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature*. 2011; 474(7351):380-384.
- Eaves LC, Wingert H, Ho HH. Screening for autism: agreement with diagnosis. *Autism*. 2006; 10(3):229-242.
- Kleinman JM, Robins DL, Ventola PE, et al. The Modified Checklist for Autism in Toddlers. *J Autism Dev Disord*. 2008; 38(5):827-839.
- Chlebowski C, Robins DL, Barton ML, Fein D. Large-scale use of the Modified Checklist for Autism in Low-Risk Toddlers. *Pediatrics*. 2013; 131(4):e1121-e1127. doi:10.1542/peds.2012-1525.
- Devlin B, Scherer SW. Genetic architecture in autism spectrum disorder. *Curr Opin Genet Dev*. 2012; 22(3):229-237.
- Roesser J. Diagnostic yield of genetic testing in children diagnosed with autism spectrum disorders at a regional referral center. *Clin Pediatr (Phila)*. 2011; 50(9):834-843.
- Rossignol DA, Frye RE. A review of research trends in physiological abnormalities in autism spectrum disorders. *Mol Psychiatry*. 2012; 17(4):389-401.
- Bauman MD, Iosif AM, Smith SE, Bregere C, Amaral DG, Patterson PH. Activation of the maternal immune system during pregnancy alters behavioral development of rhesus monkey offspring. *Biol Psychiatry*. 2014; 75(4):332-341.
- Gesundheit B, Rosenzweig JP, Naor D, et al. Immunological and autoimmune considerations of autism spectrum disorders. *J Autoimmun*. 2013; 44:1-7.
- Fu YH, Kuhl DP, Pizzuti A, et al. Variation of the CGG repeat at the fragile X site results in genetic instability: resolution of the Sherman paradox. *Cell*. 1991; 67(6):1047-1058.
- Verkerk AJ, Pieretti M, Sutcliffe JS, et al. Identification of a gene (*FMR-1*) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell*. 1991; 65(5):905-914.
- Bill BR, Geschwind DH. Genetic advances in autism: heterogeneity and convergence on shared pathways. *Curr Opin Genet Dev*. 2009; 19(3):271-278.
- Alter MD, Kharkar R, Ramsey KE, et al. Autism and increased paternal age related changes in global levels of gene expression regulation. *PLoS One*. 2011; 6(2):e16715. doi:10.1371/journal.pone.0016715.
- Glessner JT, Wang K, Cai G, et al. Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. *Nature*. 2009; 459(7246):569-573.
- Krumm N, O'Roak BJ, Shendure J, Eichler EE. A de novo convergence of autism genetics and molecular neuroscience. *Trends Neurosci*. 2014; 37(2):95-105.