Supplementary Online Content

Pramparo T, Pierce K, Lombardo MV, et al. Prediction of autism by translation and immune/inflammation coexpressed genes in infants and toddlers from pediatric community practices. *JAMA Psychiatry*. Published online March 4, 2015. doi:10.1001/jamapsychiatry.2014.3008.

This supplementary material has been provided by the authors to give readers additional information about their work.

© 2015 American Medical Association. All rights reserved.

eMethods. Supplemental Methods

SUBJECT RECRUITMENT, TRACKING AND DEVELOPMENTAL EVALUATION

Research procedures were approved by the Institutional Review Board of the University of California, San Diego. Parents of subjects underwent Informed Consent Procedures with a psychologist or study coordinator at the time of their child's enrollment.

Toddlers were recruited via the 1-Year Well-Baby Check-Up Approach, a new general population based screening approach designed to identify toddlers with an ASD around the 1st birthday or from general community sources (e.g., referred by a friend, or response to our website). In brief, the 1-Year Well-Baby Check-Up Approach utilizes a broad band screening tool, the CSBS DP IT Checklist[1] implemented at the routine first year pediatric exam. Our recent study[2], which included the participation of 137 pediatricians who implemented > 10,000 CSBS screens showed that 75% of toddlers that fail the screen at the 1st year exam have a true delay (either ASD, language delay, global developmental delay or other condition). While ASD toddlers were as young as 12 months at the time of blood sampling, all but 3 toddlers have been tracked and diagnosed using the ADOS toddler module[3] until at least age two years, an age where diagnosis of ASD is relatively stable[4-6]. Toddlers received the ADOS module that was most appropriate for their age and intellectual capacity. For the Discovery sample 64% of ASD population had an ADOST, 31% had an ADOS 1, and 5% had an ADOS 2 while for the replication sample 32% of ASD population had an ADOS T, 48% had an ADOS 1 and 20% had an ADOS 2. Only toddlers with a provisional or confirmed ASD diagnosis were included in this study. Twenty-four final diagnoses for participants older than 30 months were also confirmed with the Autism Diagnostic Interview–Revised[3]. All toddlers participated in a battery of standardized and experimental tests that included the Autism Diagnostic Observation Schedule[3], the Mullen Scales of Early Learning[7] and the Vineland Adaptive Behavior Scales[8]. Diagnoses were determined via these assessments and the Diagnostic and Statistical Manual, Fourth Edition (DSM IV-TR)[9]. Testing sessions generally lasted 4 hours and occurred across 2 separate days and the blood sample was usually taken at the end of the first day. All standardized assessments were administered by experienced Ph.D. level psychologists. In order to monitor health status, the temperature of each toddler was taken using an ear digital thermometer immediately preceding the blood draw. If temperature was higher than 99, then the blood draw was rescheduled for a different day. Parents were also asked questions regarding their child's health status such as the presence of a cold or flu, and if any illnesses were present or suspected, the blood draw was rescheduled for a different day.

GENE EXPRESSION AND DATA PROCESSING

RNA was assayed at Scripps Genomic Medicine (La Jolla, CA, USA) for labeling, hybridization, and scanning using expression BeadChips pipeline (Illumina, San Diego, CA, USA) per the manufacturer's instruction. Total RNA was extracted following standard procedures and manufacturer's instructions. In principle, LeukoLOCK disks were freed from RNA-later and Tri-reagent was used to flush out the captured lymphocyte and lyse the cells. RNA was subsequently

© 2015 American Medical Association. All rights reserved.

precipitated with ethanol and purified though washing and cartridge-based steps. The quality of mRNA samples was quantified by the RNA Integrity Number (RIN) and values of 7.0 or greater were considered acceptable[10]; all processed RNA samples passed RIN quality control. Quantification of RNA was performed using Nanodrop (Thermo Scientific, Wilmington, DE, USA). Samples were prep in 96-well plates at the concentration of 25ng/uL.

All arrays were scanned with the Illumina BeadArray Reader® and read into Illumina GenomeStudio® software (version 1.1.1). Raw data was exported from Illumina GenomeStudio® and data pre-processing was performed using the *lumi* package[11] for R (http://www.R-project.org) and Bioconductor (http://www.bioconductor.org)[12].

Several quality criteria were used to exclude low quality arrays as previously described[13,14]. In brief, low-quality arrays were those with poor signal intensity (raw intensity box plots and average signal >2 standard deviations below the mean), deviant pair-wise correlation plots, deviant cumulative distribution function plots, deviant multi-dimensional scaling plots, or poor hierarchical clustering[15]. Five samples (four ASD and one Control) were identified as low quality due to poor detection rates, different distributions and curved dot plots, and were removed prior normalization. Eighteen (18) samples had 1 replicate and all pair-wise plots of each replicas had a correlation coefficient of 0.99. Hierarchical clustering of these replicated samples showed 13 samples having with the two replicas that clustered together, therefore the B array was arbitrarily chosen for the following steps. For the remaining 5 of these replicated samples, the two replicas did not cluster together, thus the averaged gene expression levels were used in the following steps. No batch effects were identified. Raw and normalized data is deposited in Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo; GSE42133). 142 final Discovery samples/arrays (87 ASD, 55 control), and thus 142 unique subject datasets, were deemed high quality and entered the expression analysis. All 73 Replication samples/arrays (44 ASD, 29 control) were deemed high quality and used in the validation analysis of the classifier. Inter-array correlation of the Discovery dataset (IAC) was 0.983.

BrB-array filtering Tool (http://linus.nci.nih.gov/BRB-ArrayTools.html) was used to obtain a final set of genes without missing expression values. Filtering criteria were Log Intensity Variation (P>0.05) and percent missing (>50% of subjects).

DIFFERENTIAL EXPRESSION ANALYSIS ON DATA FROM ASD POST-MORTEM CORTEX

In this analysis we pulled publicly available microarray data (http://www.ncbi.nlm.nih.gov/geo/; GEO Accession ID: GSE28521) from a recent post-mortem study by Voineagu and colleagues of differences in the cortical transcriptome of ASD[16]. Differential expression (DE) analysis was done on frontal and temporal cortical tissue combined and represents the same data originally analyzed by Voineagu and colleagues in their differential expression analysis. Our DE analysis was implemented in Matlab R2012b and utilized functions from the Bioinformatics toolbox for the main DE analysis steps (i.e. mattest.m, mafdr.m, and clustergram.m). The function mattest.m was used to run permutation t-tests (10,000 permutations) on every gene in the microarray

© 2015 American Medical Association. All rights reserved.

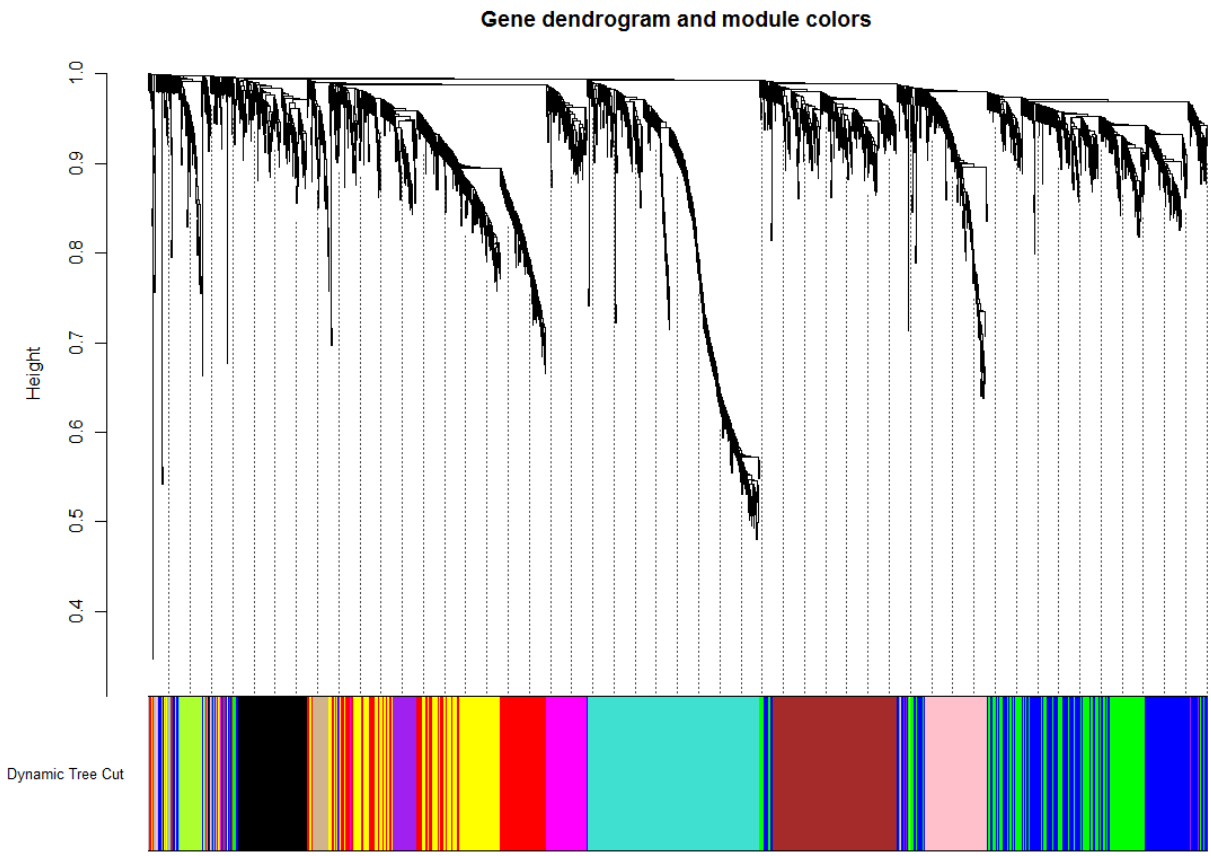
dataset. The function mafdr.m was then use to compute the FDR corrected p-values (using Storey's method)[17] at a threshold of $q<0.01$.

We found 1272 downregulated and 783 upregulated genes in ASD cortex. eFigure 2A below shows a two-way unsupervised clustering (using the function clustergram.m) on the final set of downregulated and upregulated DE genes. MetaCore functional enrichment analysis showed that the 783 genes that are upregulated in ASD cortex are highly enriched in processes we also observe as being prominent in our blood-expression classifier (see eFigure 2B), such as Translation_Translation_initiation, Translation_Elongation-Termination, Translation_Elongation-Termination_test, Immune_response_Phagocytosis, and Cell_cycle_G1- S_Growth_factor_regulation. In contrast, the 1272 genes downegulated in ASD are highly enriched for synaptic involvement and other neurophysiological processes, which is congruent with Voineagu and colleagues earlier report using GO enrichment analysis (see eFigure 2C), but only one of these processes (Immune_response_TCR_signaling) is prominent in our classifier.

> This analysis underscores some very important points regarding what we have found in our blood-expression signature.

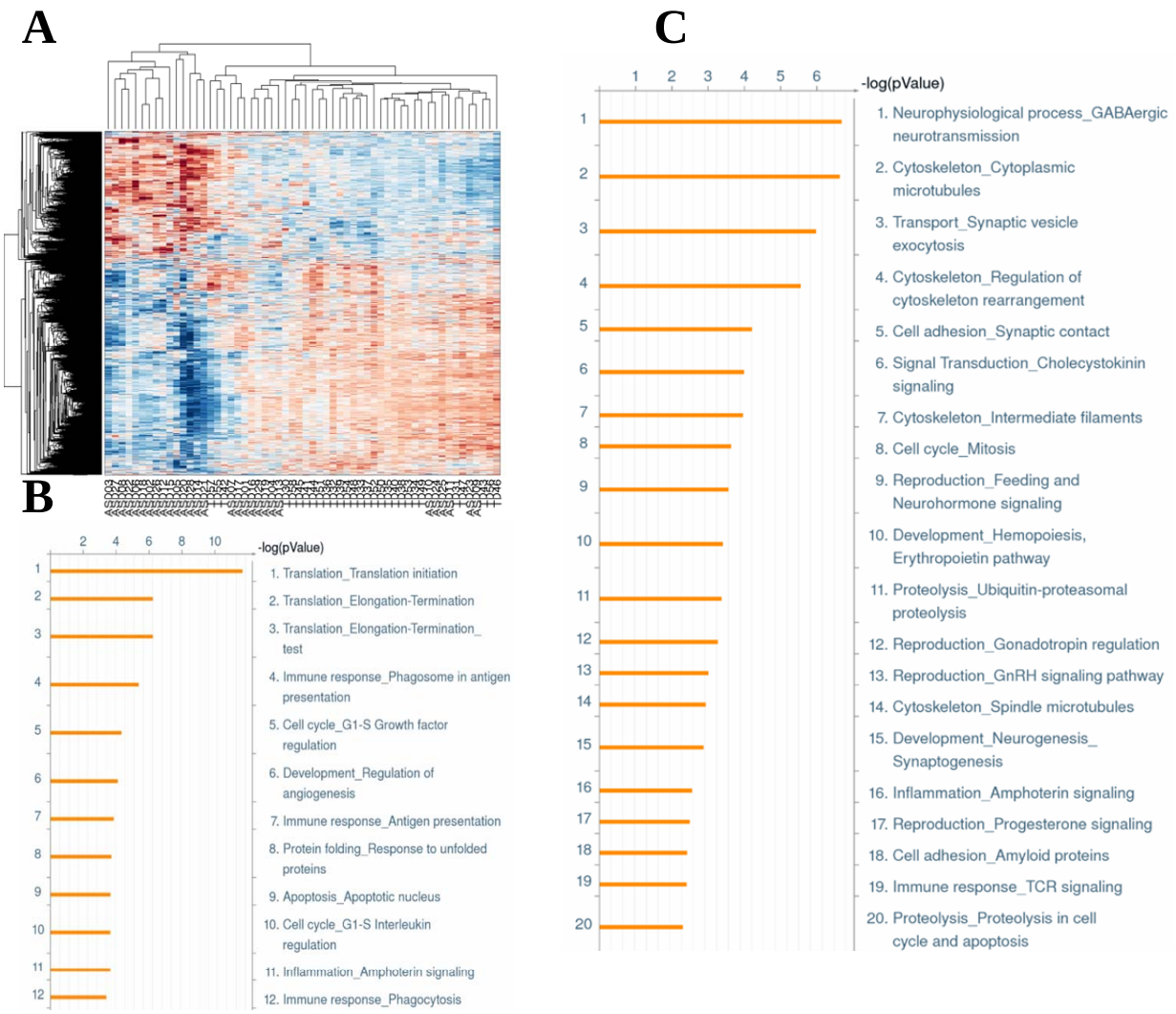
First, it shows that much of the signature in blood appears to overlap with genes that are known to be over-expressed (i.e. upregulated) in ASD cortex, and that enrichment analysis is largely congruent across blood and brain in this manner. Second, our blood-expression classifier seems not as sensitive to processes known to be downregulated in ASD cortex that have to do with processes that are highly specific to the brain and brain tissue (e.g., synaptic processes). This largely confirms the caveat that while other studies report a very significant contribution of abnormalities occurring at the synapse, expression in blood may not be as sensitive to picking up on such brain-specific signals, but is very sensitive to system-wide processes that do have much relevance for the brain in ASD (e.g., translation, immune/inflammation, cell cycle processes).

© 2015 American Medical Association. All rights reserved.



eFigure 1. WGCNA Analysis Across ASD and Control Toddlers Using the Differentially Expressed Genes. Co-expression modules are generated and color-coded. Each vertical line correspond to a gene, and genes with similar patterns are clustered into modules.
Modules are herein called by the assigned WGCNA default colors. Module eigengenes are computed for each subject and each module.

© 2015 American Medical Association. All rights reserved.



eFigure 2. Differentially Expressed (DE) Genes in ASD Postmortem Cortical Tissue and Enrichment for Genes Downregulated or Upregulated in ASD. Panel A shows a two-way unsupervised hierarchical clustering of differentially expressed genes, with genes along the rows and individual subjects along the columns. Genes that are upregulated in ASD can be seen in the upper rows of the heatmap, with most of the ASD subjects showing warm values and most control subjects (TD) showing cooler values. Downregulated genes can be seen in the lower rows of the heatmap with cooler values for most ASD subjects and warmer values for most control subjects. Panel B shows the MetaCore enrichment analysis for genes upregulated in ASD. This enrichment analysis is very similar in key processes that also appear in our classifier. Panel C shows the MetaCore enrichment analysis for genes downregulated in ASD.

© 2015 American Medical Association. All rights reserved.

eTable. CNV Analysis of Misclassified ASD Subjects

SubjectID	CNV location (hg18)	Size (bp)	DEL/DUP	Genes involved
X3F5T	chr6:169182781-	540,264	DUP	AK055570,BX648586,THBS2,WDR27
X3F5T	chr7:147713357-	106,109	DUP	**CNTNAP2**,LOC392145
M8K5X	chr20:47589174-47611381	22,207	DEL	PTGIS
Y2B4P	chr15:21744675-21778678	34,003	DEL	Intergenic (NDN, AK124131)
J3L5W	chr1:231796069-	17,644	DEL	Intergenic (KIAA1804, KCNK1)
L5S3Z	chr1:242582713-	18,934	DEL	C1orf100
X2H3X	chr12:72374075-72391780	17,705	DUP	Intergenic (TRHDE, BC061638)
J3L5W	chr14:19754117-19825127	71,010	DEL	OR11H4,OR11H6
S3D7F	chr5:12578748-12906570	327,822	DEL	AY328033,AY330599
Z3W7W	chr6:132884089-	22,152	DEL	TAAR9

DEL=heterozygous deletion, DUP=duplication. Reference genome hg18.

© 2015 American Medical Association. All rights reserved.

eREFERENCES

1. Wetherby, A.M., Allen, L., Cleary, J., Kublin, K. & Goldstein, H. Validity and reliability of the communication and symbolic behavior scales developmental profile with very young children. *J Speech Lang Hear Res* **45**, 1202-18 (2002).
2. Pierce, K. *et al.* Detecting, studying, and treating autism early: the one-year well-baby check-up approach. *J Pediatr* **159**, 458-465 e1-6 (2011).
3. Luyster, R. *et al.* The Autism Diagnostic Observation Schedule-toddler module: a new module of a standardized diagnostic measure for autism spectrum disorders. *Journal of Autism and Developmental Disorders* **39**, 1305- 1320 (2009).
4. Chawarska, K., Klin, A., Paul, R., Macari, S. & Volkmar, F. A prospective study of toddlers with ASD: short-term diagnostic and cognitive outcomes. *J Child Psychol Psychiatry* **50**, 1235-45 (2009).
5. Cox, A. *et al.* Autism spectrum disorders at 20 and 42 months of age: stability of clinical and ADI-R diagnosis. *J Child Psychol Psychiatry* **40**, 719-32 (1999).
6. Kleinman, J.M. *et al.* Diagnostic stability in very young children with autism spectrum disorders. *J Autism Dev Disord* **38**, 606-15 (2008).
7. Mullen, E.M. *Mullen Scales of Early Learning*, (American Guidance Service Inc., MN, 1995).
8. Sparrow, S., Cicchetti DV, Balla DA. Vineland Adaptive Behavior Scales. Second Edition. Survey Forms Manual.
   Pearson Assessments (2005).
9. Association, A.P. Diagnostic and Statistical Manual of Mental Disorders. Fourth Edition,. *American Psychiatric Association* (2000).
10. Schroeder, A. *et al.* The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC Mol Biol* **7**, 3 (2006).
11. Du, P., Kibbe, W.A. & Lin, S.M. lumi: a pipeline for processing Illumina microarray. *Bioinformatics (Oxford, England)* **24**, 1547-8 (2008).
12. Gentleman, R.C. *et al.* Bioconductor: open software development for computational biology and bioinformatics.
    Genome Biology **5**, R80 (2004).
13. Chow, M.L. *et al.* Genome-wide expression assay comparison across frozen and fixed postmortem brain tissue samples. *BMC genomics* **12**, 449 (2011).
14. Chow, M.L. *et al.* Age-dependent brain gene expression and copy number anomalies in autism suggest distinct pathological processes at young versus mature ages. *PLoS Genet* **8**, e1002592 (2012).
15. Oldham, M.C. *et al.* Functional organization of the transcriptome in human brain. *Nature Neuroscience* **11**, 1271- 1282 (2008).
16. Voineagu, I. *et al.* Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature* **474**, 380-4 (2011).
17. Storey, J.D. A direct approach to false discovery rates. *J. R. Statist. Soc. B* **64**, 479-498 (2002).

© 2015 American Medical Association. All rights reserved.