NIH Public Access

Author Manuscript

J Am Acad Child Adolesc Psychiatry. Author manuscript; available in PMC 2013 September 01.

Published in final edited form as:

J Am Acad Child Adolesc Psychiatry. 2012 September ; 51(9): 934–44.e2. doi:10.1016/j.jaac.2012.07.007.

Blood-Based Gene Expression Signatures of Autistic Infants and Toddlers:

RH: Blood-Based Autism Biomarkers

Stephen J. Glatt, PhD, **Ming T. Tsuang, MD**, **Mary Winn, PhD**, **Sharon D. Chandler, PhD**, **Melanie Collins**, **Linda Lopez, ANS**, **Melanie Weinfeld, PhD**, **Cindy Carter, PhD**, **Nicholas Schork, PhD**, **Karen Pierce, PhD**, and **Eric Courchesne, PhD**

Dr. Glatt is with the Psychiatric Genetic Epidemiology and Neurobiology Laboratory (PsychGENe Lab), Medical Genetics Research Center, State University of New York (SUNY) Upstate Medical University. Dr. Tsuang is with Harvard Institute of Psychiatric Epidemiology and Genetics, Harvard School of Public Health, and Harvard Medical School; the Center for Behavioral Genomics, University of California–San Diego (UCSD); Veterans Affairs San Diego Healthcare System; and the Institute of Genomic Medicine, UCSD. Dr. Winn is with UCSD and the Scripps Translational Science Institute. Drs. Chandler and Collins are with the Center for Behavioral Genomics, UCSD. Drs. Lopez, Weinfeld, Carter, Pierce, and Courchesne are with the Autism Center of Excellences, UCSD. Dr. Schork is with the Scripps Translational Science Institute

Abstract

Objective—Autism spectrum disorders (**ASDs**) are highly heritable neurodevelopmental disorders that onset clinically during the first years of life. ASD-risk biomarkers expressed early in life could significantly impact diagnosis and treatment, but no transcriptome-wide biomarker classifiers derived from fresh blood samples from children with autism have yet emerged.

Method—Using a community-based, prospective, longitudinal method, we identified 60 infants and toddlers at-risk for ASDs (autistic disorder and pervasive developmental disorder), 34 at-risk for language delay (**LD**), 17 at-risk for global developmental delay (**DD**), and 68 typically developing (**TD**) comparison children. Diagnoses were confirmed *via* longitudinal follow-up. Each child's mRNA expression profile in peripheral blood mononuclear cells (**PBMCs**) was determined by microarray.

Results—Potential ASD biomarkers were discovered in one half of the sample and used to build a classifier with high diagnostic accuracy in the remaining half of the sample.

Conclusions—The mRNA expression abnormalities reliably observed in PBMCs, which are safely and easily assayed in babies, offer the first potential peripheral blood-based early biomarker panel of risk for autism in infants and toddlers. Future work should verify these biomarkers and evaluate if they may also serve as indirect indices of deviant molecular neural mechanisms in autism.

© 2013 American Academy of Child & Adolescent Psychiatry. Published by Elsevier Inc. All rights reserved

Correspondence to Stephen J. Glatt, Ph.D., SUNY Upstate Medical University, 750 East Adams Street, Syracuse, NY, 13210, U.S.A.; glatts@upstate.edu.

Disclosure: Drs. Tsuang, Winn, Chandler, Weinfeld, Carter, Schork, Pierce, and Courchesne, and Ms. Collins and Ms. Lopez report no biomedical financial interests or potential conflicts of interest.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Keywords

autism; biomarker; classifier; microarray; support vector machine

INTRODUCTION

Autism spectrum disorders (**ASD**s), such as autistic disorder (**AD)** and pervasive developmental disorder not otherwise specified (**PDD**), are among the most prevalent neurodevelopmental disorders, collectively affecting almost 1% of the U.S. population.[1] These are also among the most severe and debilitating disorders of early brain development, typically emerging between infancy and early childhood, and commonly imposing lifelong disability on affected children as well as considerable strain on their caretakers and society at large. One way to combat ASDs would be to discover biological signatures—or "biomarkers"—of early risk for ASDs which could be exploited for early detection and treatment and, thereby, improve long-term clinical outcomes.[2]

Most prior transcriptome-wide studies of gene-expression in ASDs have utilized *postmortem* brain tissue[3, 4] or transformed lymphoblastoid cell lines.[5–8] Both approaches are quite useful and revealing, often observing dysregulation of viable candidate pathways in autism, but both have limitations, such as the inability to fully account for agonal factors or the effects of viral transformation on the observed gene-expression differences in autism. In contrast, Kuwano *et al.*[9] sought autism biomarkers in peripheral leukocytes of young adults with autism, observing dysregulation of 16 genes related to cell morphology, cellular assembly and organization, and nerve system development and function. While informative regarding persistent gene-expression abnormalities in peripheral blood in individuals with ASDs (and perhaps reflecting the central nervous system pathology accompanying the disorder), this study cannot shed light on the peripheral-blood transcriptomic events accompanying the emergence of autism in its earliest developmental stages of infancy or toddlerhood.

Here we report the cross-sectional results of whole-transcriptome expression profiling in freshly drawn peripheral blood mononuclear cells (**PBMC**s) from Wave I of data-collection in a longitudinal study of very young children (12–46 months) with or at-risk for a variety of neurodevelopmental disorders including ASDs, language delay (**LD**), or global developmental delay (**DD**), as well as children consistently judged to be typically developing (**TD**) and children initially judged as potentially at-risk but, upon further evaluation, found to be developing typically (type-I errors; **TIE**s). Our analyses were designed to reveal in peripheral blood a biomarker signature for ASDs that might also reveal important biological differences between TD children, children with delays not on the autism spectrum, and children at-risk for or having an ASD.

METHOD

Subjects

Protections—The Institutional Review Boards of the University of California, San Diego and SUNY Upstate Medical University approved all procedures. After referral, all minor subjects assented to the study procedures, and one or both parents or legal guardians of each subject provided written informed consent for their child to participate.

Ascertainment—Participants were obtained *via* a general population-based screening method called the One-Year Well-Baby Check-Up Approach, which engages community pediatricians to screen for autism using the Communication and Symbolic Behavior Scales

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Developmental Profile (**CSBS DP**) Infant-Toddler questionnaire[10, 11] during regular checkups.[12] Over 135 pediatricians in San Diego County agreed to screen all consecutive infants and toddlers seen for routine visits. All subjects failing the CSBS DP at their pediatrician's office were provisionally identified as at-risk for one of the disorders of interest and then referred to our Center for further evaluation. Subjects who passed the CSBS were identified as TD, and pediatricians randomly referred such subjects to our Center for further evaluation as unaffected comparison subjects. Using this approach, toddlers as young as 12-months were recruited and tracked every six months until at least their third birthday, thus allowing for the prospective study of autism beginning at 12 months.

Diagnoses—ASDs were initially diagnosed based on the clinical judgment of a Ph.D.-level psychologist with over 10 years of experience in autism using the three initial modules of the Autism Diagnostic Observation Schedule (**ADOS**)[13] and the revised algorithms described by Gotham *et al.*[14] While several ASD subjects were only one year old at the time of first blood sampling, all but one have been tracked and diagnosed with an ASD using the toddler module of the ADOS[15] at age two or beyond, when the diagnosis of autism can be made reliably. Final diagnoses for participants with an ASD older than 30 months (including some of the youngest subjects who had previously been diagnosed with or at-risk for an ASD based on the ADOS) were confirmed with the Autism Diagnostic Interview–Revised (**ADI-R**).[15] Subjects who failed the CSBS DP Infant-Toddler questionnaire but who passed the ADOS were diagnosed with LD if one or both of the language subtest scores of the Mullen Scales of Early Learning were more than one standard deviation lower than expected for that age (T-score <40). Since a significant percentage of individuals with ASDs exhibit language deficits not restricted to the pragmatic component of language,[16–19] this approach to diagnosing LD may not be highly specific; however, this criterion paired with ADOS evaluations not consistent with an ASD is sufficient to establish that these subjects have some language delay and are not on the autistic spectrum, making them suitable as an affected, non-autistic comparison sample for establishing the specificity of our ASD biomarkers. Subjects who failed the CSBS DP Infant-Toddler questionnaire but who passed the ADOS were diagnosed with global DD if scores were more than one standard deviation lower than expected on three or more subtests of the Mullen Scales and the overall developmental quotient was more than one standard deviation lower than expected (<85). Subjects who failed the CSBS DP Infant-Toddler questionnaire but who passed the ADOS and who did not deviate from norms on any of the Mullen Scales or the developmental quotient were judged to be developing typically and identified as TIEs.

Laboratory Methods

From each subject, 4ml of venous blood was collected into EDTA-coated collection tubes and immediately transferred to an RNase-free laboratory where all subsequent procedures took place, including RNA isolation and storage at −80°C within six hours. Methods of mRNA extraction, stabilization, isolation, storage, quantitation, reverse transcription, and microarray hybridization, as well as methods of microarray scanning and data import, normalization, and transformation, were consistent with our prior work,[20] (see Supplemental Method, available online).

Quality Control and Final Sample Constitution

The quality of mRNA samples was quantified by the RNA Integrity Number (**RIN**) and, according to convention,[21] values of 6.0 or greater were deemed acceptable; values observed in our sample ranged from 9.0–10.0. A total of 339 samples selected for analysis in Wave I had acceptable levels of mRNA quantity, purity, and quality. Once microarray data were generated, the first three principal components of global gene expression patterns in each

individual were visualized. Subjects whose values on each of these three principal components were beyond four standard deviations from the grand mean (n=5) were removed from further analyses, as this type of deviation indicates that typical patterns of global gene co-expression and correlation have been disrupted (by either a biological process or, more likely in our experience, a laboratory technical problem). Eighteen additional subjects were removed due to low signal intensity as observed by box-plot and evidenced by an average signal more than two standard deviations below the mean, which likely indicates generally low hybridization efficiency and unreliability of the data from those arrays. Of the remaining 316 quality-assured samples, 179 were first-timepoint samples from a child proband in one of the six diagnostic groups described in Table 1. The remaining 137 samples were from parents or siblings of probands, or from probands evaluated at a second timepoint, and so were not formally analyzed for the present report. There was no significant overall sex difference between the six diagnostic groups ($\chi^2_{[5]}$=10.61, p=0.060), but they differed significantly in age ($F[5,173]$=13.77, p<0.0001), reflecting that the LD and TIE groups were significantly younger than the AD, DD, PDD, and TD groups (Scheffe test, all p<0.05). This age difference should be put into context however, as the full age range for our entire sample is only 34 months, which is still relatively narrow for a study of early autism; nevertheless, all subsequent statistical analyses accounted for age as a covariate. As expected, AD and PDD groups differed significantly from both the TD and the TIE control groups on all clinical scale scores, including the Mullen Early Learning Composite Score, Vineland Adaptive Behavior Scales Standard Score, Autism Diagnostic Observation Schedule–Toddler (ADOS-T) Social Affect Score, ADOS-T Restricted/Repetitive Score, and ADOS-T Total Score. The AD and PDD groups also differed from each other on the ADOS-T Social Affect and Total Scores, with the AD group showing overall greater symptom severity. The majority of subjects in the full sample and within each diagnostic group was of Caucasian ancestry, with no other ancestral group (Asian, Pacific Islander, African-American, American Indian, or Mixed) individually comprising more than 10% of any diagnostic group, and there were no differences in overall ancestral composition between any of the diagnostic groups (Pearson $\chi^2_{(25)}$=21.516; p=0.664). To increase sample size and inferential power, and maintain consistency with the conceptualization of AD and PDD as disorders on the same spectrum, we pooled these samples together into one ASD group for all subsequent analyses; likewise, we pooled both typically developing groups (TD and TIE) into one unaffected control (**CNT**) group for all subsequent analyses, acknowledging that other interesting combinations and comparisons could be tested.

Microarray Data Analyses

The ASD sample was split into two independent subgroups: a discovery sample and a replication sample. The two subsamples were drawn to be equivalent in sample size, age, sex, and diagnostic composition (AD *vs.* PDD). The CNT sample was similarly split into discovery and replication samples drawn to be equivalent in size, age, sex, and diagnostic composition (TD *vs.* TIE). The basic analytic model for identifying candidate biomarkers in the discovery samples was an analysis of covariance (**ANCOVA**) with each gene's expression intensity value as the dependent measure, diagnostic group (ASD *vs.* CNT) and sex as fixed between-subjects factors, and age in months as a continuous covariate. Genes whose expression levels were influenced by at least a nominally significant main effect of diagnostic group (p<0.05) and whose *log*2 fold-change in the ASD group was |1.2| were identified as putative candidate ASD biomarkers and advanced to the next phase of analysis. These fairly liberal criteria were used to cast a wide net to catch all potentially informative genes, while false-positives would be pared off the putative biomarker set by subsequent model-building, -optimization, and replication steps. These genes were then used to build a support vector machine (**SVM**) classifier of the same discovery sample in which they were initially discovered. The SVM was configured as cost-based, with costs varied from 1 to

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

1001 in intervals of 100. The tolerance (termination criterion) of the SVM was set at 0.001. The kernel for the SVM was a radial basis function, with gamma equal to the inverse of the number of evaluated markers. The optimal (*i.e.*, most accurate and parsimonious) SVM was derived by shrinking centroids (which prunes highly correlated or redundant features) and 10-fold cross-validation. The optimal SVM classifier was then tested for classification accuracy in the fully independent ASD and CNT replication samples, and then re-evaluated in the presence of the LD and DD samples as affected but non-autistic comparators. The list of genes comprising the optimal classifier was subjected to the Database for Annotation, Visualization, and Integrated Discovery (**DAVID**)[22] algorithm to determine if it was enriched with genes having particular functional, ontological, or structural annotations. Fold-enrichments of such annotations were only declared statistically significant if they surpassed a Bonferroni-corrected p-value threshold of α=0.05/the number of terms evaluated in a particular category.

RESULTS

Discovery Samples

Putative Biomarker Identification by ANCOVA—We found 2540 probes whose expression levels were at least nominally significantly different ($0.05>p>3.0e^{-6}$) between the ASD and CNT groups. When limiting this probe-set to only those that showed significant dysregulation in ASD ($p<0.05$) and a *log*2 fold-change |1.2|, 154 probes remained (Table 2). The *log*2 fold-change in these probes ranged from 2.1-fold up-regulation to 1.7-fold down-regulation, corresponding to actual changes of 4.3-fold up-regulation and 3.2-fold down-regulation, respectively.

Construction of Optimal SVM Classifier—The 154 probes meeting our significance and fold-change criteria were utilized to construct, optimize, and test the classification accuracy of various SVMs. The most accurate SVM (cost 901, gamma 0.001) utilized the expression intensities of 48 probes (Table 2) to correctly classify 71% of ASD and CNT subjects across ten subsets of the discovery sample into their appropriate diagnostic categories. Twenty-two of 30 ASD subjects were correctly classified by this SVM as having an ASD, while 23 of 34 CNT subjects were correctly classified as CNT subjects. The corresponding sensitivity and specificity of this SVM after ten-fold cross-validation was 0.73 and 0.68, respectively.

Replication Samples

Evaluation of Optimal SVM Classifier—The identical 48-probe SVM classifier optimized in the discovery sample performed almost perfectly in the replication sample, with an accuracy of 91%. Just 2 of 30 ASD subjects were incorrectly identified as CNT subjects, while only 4 of 34 CNT subjects were misclassified as ASD subjects. The corresponding sensitivity and specificity of this SVM was 0.93 and 0.88, respectively. The area under the receiver-operating characteristic (ROC) curve generated from these classifications was 0.91.

When the optimized 48-probe SVM diagnostic classifier of ASDs was deployed to classify the ASD and CNT replication samples in the presence of LD and DD samples, its accuracy was only slightly attenuated to 87%. Only 2 of 30 ASD subjects in the replication sample were misclassified (1 as CNT and 1 as LD), and only 8 of 36 individuals classified by the SVM as ASD group members were not actually diagnosed as such (including 4 CNT subjects, 1 DD subject, and 3 LD subjects). Though not designed or optimized for this purpose, the 48-probe SVM classifier of ASD also fared quite well in distinguishing LD and

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

DD subjects from CNT subjects and from each other, with only 8 errors made among the total of 80 subjects assigned to these categories.

The list of 48 probes comprising the optimal SVM classifier of ASDs was most significantly enriched with genes related to immune responses in general (nine genes: *APOBEC3F*, *APOBEC3G*, *FCGR1B*, *FCGR1C*, *GBP1*, *GBP4*, *GBP5*, *GCH1*, and *TAP1*; six-fold enrichment compared to chance expectation; Bonferroni-corrected p=0.014), and genes of the hemoglobin complex (four genes: *AHSP*, *HBD*, *HBG1*, and *HBG2*; 142-fold enrichment; Bonferroni-corrected p=1.8E^{-4}) and genes with guanine- or guanylate-binding affinity (three genes: *GBP1*, *GBP4*, and *GBP5*; 140-fold enrichment; Bonferroni-corrected p=3.7E^{-3}) in particular. Although no known canonical pathways were disproportionally represented among the 48 genes, some links between the genes along various biological pathways were observed through Ingenuity Pathway Analysis (see Figure S1, available online).

DISCUSSION

In this study we identified a blood-based gene expression profile that reliably identified subjects with ASDs near the age of first clinical signs. In fact, somewhat surprisingly, the optimal classifier attained nominally higher accuracy (though not statistically significantly higher) in the test sample than in the training sample. More typically, a classifier will perform best in the sample in which it was trained, and a drop-off in performance will be encountered in subsequent test sets. In the present case, the apparent difference in classification accuracy may stem from the fact that the accuracy in the training sample actually represents the average of ten smaller cross-validation samples (each with just 4–5 cases and 4–5 comparison subjects), whereas the value in the test set is derived from measurement in the full test sample without additional cross-validation. Or the spike in accuracy may represent a random fluctuation, and additional test samples may very well show a decrement in accuracy, which would not be unexpected. In either event, further analyses and outside replications of the optimal gene-expression-based classifier of ASDs are needed.

This promising work, while far from being definitive, gives further proof to the recently emerging principle that peripheral blood is a potentially useful source of biomarkers for disorders of the brain and other inaccessible tissues.[23] The optimal ASD classifier panel of 48 mRNA biomarkers, which was derived to maximize differences between ASD subjects and those in the two unaffected control groups (TD and TIE), might ultimately have clinical utility in the context of high-risk screening where a heightened risk of autism is feared or expected (*i.e.*, in multiply affected families). However, this same model was not necessarily expected to provide maximal classification accuracy for distinguishing ASD subjects from those with other developmental delays or disorders, such as the LD and DD subjects in our sample. Nevertheless, when we performed these comparisons, the 48-probe model of ASDs still retained a very high degree of classification accuracy. This model, in contrast, may ultimately have utility (and could be further optimized) for a general population screen for ASDs at an early age, or in the form of a test for autism among children who show at least some early warning signs of developmental delay, whether or not those warning signs ultimately manifest as a diagnosable disorder (*e.g.*, ASDs, DD, or LD) or not (*e.g.*, TIE).

Since we first demonstrated the potential utility of transcriptome profiling in peripheral blood cells as a source of biomarkers for mental illnesses in 2005,[24] we and others have sought to identify those select biomarkers that also might shed light on the pathophysiology underlying such disorders.[20, 25] Further work will be needed to establish if these dysregulated biomarkers are truly indicative of the early developmental pathogenesis of

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

autism in the brain, and if the dysregulation of these genes is governed by one or more expression quantitative trait loci; however, neither is a prerequisite of an effective biomarker. Yet, one of the 48 biomarkers in our optimized SVM (*IFI16*, which encodes interferon, gamma inducible 16) was previously found to be dysregulated in *postmortem* temporal cortex of subjects with autism.[26] In turn, although some of the identified biomarkers may highlight potential pathophysiological processes, our work is no replacement for studies utilizing animal models, human *postmortem* brain tissue, induced pluripotent stem cells, and neuroimaging studies of children living with autism.

Our analyses of gene-expression data at just one early timepoint in the lives of actively developing children preclude us from determining if the identified biomarkers of ASDs are or will be stable over time. A recent study of longitudinal gene expression in healthy individuals,[27] as well as earlier small-scale studies,[28–31] found remarkable temporal stability in the expression of most genes within individuals, with only 1–2% of genes showing significant variability over periods on the order of one month or longer; however, these studies have not routinely examined children, nor individuals affected with neurodevelopmental disorders such as ASDs. Thus, the stability of the particular biomarker signatures we identified remains an empirical question to be answered in due course. Toward that objective, we continue to track the subjects reported on here, as well as ascertain new subjects in this study, which will provide an invaluable source of replication since other similar sample resources are not yet available elsewhere. As we are tracking these subjects longitudinally, we can observe how these putative biomarker profiles change over time as symptoms also change naturally or in response to intensive behavioral treatment. Further evaluation of LD and DD subjects will also shed light on the unique biomarker signatures of these important conditions. Perhaps most importantly, the peripheral blood gene expression data (and linked genetic polymorphism and structural and functional neuroimaging data) generated in our sample of children with a broad range of clinical symptoms both within and beyond the autistic spectrum could be harnessed for the purpose of “reverse engineering” a new nosology of childhood neurodevelopmental disorders based on clusters of biological homogeneity rather than clinical symptoms which may reflect common endpoints of diverse etiologies.

The putative ASD biomarkers found here were initially identified using an ANCOVA that accounted for the age differences in the groups, and the marker set advancing to the stage of classifier-construction was filtered for genes showing a significant age effect. Thus, the age difference between diagnostic groups is not expected to be a major driver of classification accuracy. We recognize, however, that in real-world applications a child's age—in conjunction with a particular gene-expression profile—may be critically important for making an ASD diagnosis. Therefore, with the addition of more subjects in this study across a range of ages, we continue working to refine distinct ASD-biomarker signatures for children of different ages, and to incorporate age itself (along with the expression levels of both age-dependent and -independent genes) as a factor influencing the performance of the diagnostic classifier.

Much work remains to be done, but even at this stage the currently characterized biomarkers —if replicated—could be harnessed by efforts to expedite and standardize the process of primary and differential diagnosis, which presently involves considerable time, effort, and uncertainty. Biomarkers such as these, if deployed in a clinical testing environment, should allow for earlier identification of affected individuals, in turn hastening their receipt of effective (perhaps personalized) treatment and improving prognoses.[2] Furthermore, biomarkers could form the basis for early intervention and prevention efforts targeting at-risk infants and toddlers, which might significantly improve clinical outcomes for affected infants.[32] The identified biomarkers may, in turn, suggest “treatable targets” for future

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

waves of medicinal chemistry and drug-discovery research. Collectively, these advances would translate into an enormous improvement in global public health.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported in part by U.S. National Institutes of Health (NIH) grants P50MH081755, R01MH080134, R21MH075027, U19AG023122, R01MH078151, N01MH22005, U01DA024417, UL1RR025774, RC2DA029475, R01AG031224, U54NS056883, and P50MH081755-0003, the Price Foundation, Scripps Genomic Medicine, the Sidney R. Baer, Jr. Foundation, and the Brain and Behavior Research Foundation (NARSAD).

The authors wish to thank the children and family members who participated in this study and Cheryl A. Roe, Timothy Wong, and Justin Zelenka of SUNY Upstate Medical University for their outstanding technical assistance.

Dr. Glatt has received grant or research support from NIH and the Brain and Behavior Research Foundation (NARSAD). He has served as a consultant to SynapDx, Corp. and has received royalties from Oxford University Press. He holds U.S. Patent #7,879,548.

REFERENCES

1. Autism and Developmental Disabilities Monitoring Network Surveillance Year 2006 Principal Investigators. Prevalence of autism spectrum disorders - Autism and Developmental Disabilities Monitoring Network, United States, 2006. MMWR Surveill Summ. Dec 18; 2009 58(10):1–20.
2. Dawson G. Recent advances in research on early detection, causes, biology, and treatment of autism spectrum disorders. Curr Opin Neurol. Apr; 2010 23(2):95–96. [PubMed: 20216345]
3. Purcell AE, Jeon OH, Zimmerman AW, Blue ME, Pevsner J. Postmortem brain abnormalities of the glutamate neurotransmitter system in autism. Neurology. Nov 13; 2001 57(9):1618–1628. [PubMed: 11706102]
4. 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. Mar.2012 8(3):e1002592. [PubMed: 22457638]
5. Hu VW, Nguyen A, Kim KS, et al. Gene expression profiling of lymphoblasts from autistic and nonaffected sib pairs: altered pathways in neuronal development and steroid biosynthesis. PLoS One. 2009; 4(6):e5775. [PubMed: 19492049]
6. Baron CA, Liu SY, Hicks C, Gregg JP. Utilization of lymphoblastoid cell lines as a system for the molecular modeling of autism. J Autism Dev Disord. Nov; 2006 36(8):973–982. [PubMed: 16845580]
7. Nishimura Y, Martin CL, Vazquez-Lopez A, et al. Genome-wide expression profiling of lymphoblastoid cell lines distinguishes different forms of autism and reveals shared pathways. Hum Mol Genet. Jul 15; 2007 16(14):1682–1698. [PubMed: 17519220]
8. Ghahramani Seno MM, Hu P, Gwadry FG, et al. Gene and miRNA expression profiles in autism spectrum disorders. Brain Res. Mar 22.2011 1380:85–97. [PubMed: 20868653]
9. Kuwano Y, Kamio Y, Kawai T, et al. Autism-associated gene expression in peripheral leucocytes commonly observed between subjects with autism and healthy women having autistic children. PLoS One. 2011; 6(9):e24723. [PubMed: 21935445]
10. Wetherby, AM.; Prizant, BM. Communication and Symbolic Behavior Scales: Developmental Profile. 1st normed ed. Paul H Brookes Publishing; Baltimore, MD, US: 2002.
11. Wetherby AM, 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. Dec; 2002 45(6):1202–1218. [PubMed: 12546488]
12. Pierce K, Carter C, Weinfeld M, et al. Detecting, Studying, and Treating Autism Early: The One-Year Well-Baby Check-Up Approach. J Pediatr. Apr 19.2011

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

13. Lord, C.; Rutter, M.; DiLavore, P.; Risi, S. Autism Diagnostic Observation Schedule (ADOS) manual. Western Psychological Services; Los Angeles, CA: 2001.
14. Gotham K, Risi S, Pickles A, Lord C. The Autism Diagnostic Observation Schedule: revised algorithms for improved diagnostic validity. J Autism Dev Disord. Apr; 2007 37(4):613–627. [PubMed: 17180459]
15. Luyster R, Gotham K, Guthrie W, et al. The Autism Diagnostic Observation Schedule-toddler module: a new module of a standardized diagnostic measure for autism spectrum disorders. J Autism Dev Disord. Sep; 2009 39(9):1305–1320. [PubMed: 19415479]
16. Tager-Flusberg H, Cooper J. Present and future possibilities for defining a phenotype for specific language impairment. J Speech Lang Hear Res. Oct; 1999 42(5):1275–1278. [PubMed: 10515521]
17. Rapin I, Dunn M. Update on the language disorders of individuals on the autistic spectrum. Brain Dev. Apr; 2003 25(3):166–172. [PubMed: 12689694]
18. Tager-Flusberg H, Joseph RM. Identifying neurocognitive phenotypes in autism. Philos Trans R Soc Lond B Biol Sci. Feb 28; 2003 358(1430):303–314. [PubMed: 12639328]
19. Tager-Flusberg H, Skwerer DP, Joseph RM. Model syndromes for investigating social cognitive and affective neuroscience: a comparison of Autism and Williams syndrome. Soc Cogn Affect Neurosci. Dec; 2006 1(3):175–182. [PubMed: 18985104]
20. Glatt SJ, Everall IP, Kremen WS, et al. Comparative gene expression analysis of blood and brain provides concurrent validation of SELENBP1 up-regulation in schizophrenia. Proceedings of the National Academy of Sciences of the United States of America. 2005; 102(43):15533–15538. [PubMed: 16223876]
21. Schroeder A, Mueller O, Stocker S, et al. The RIN: an RNA integrity number for assigning integrity values to RNA measurements. BMC Molecular Biology. 2006; 7:3. [PubMed: 16448564]
22. Dennis G Jr, Sherman BT, Hosack DA, et al. DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biology. 2003; 4(5):P3. [PubMed: 12734009]
23. Kohane IS, Valtchinov VI. Quantifying the white blood cell transcriptome as an accessible window to the multiorgan transcriptome. Bioinformatics. Feb 15; 2012 28(4):538–545. [PubMed: 22219206]
24. Tsuang MT, Nossova N, Yager T, et al. Assessing the validity of blood-based gene expression profiles for the classification of schizophrenia and bipolar disorder: a preliminary report. Am J Med Genet B Neuropsychiatr Genet. Feb 5; 2005 133B(1):1–5. [PubMed: 15645418]
25. Sullivan PF, Fan C, Perou CM. Evaluating the comparability of gene expression in blood and brain. Am J Med Genet B Neuropsychiatr Genet. Apr 5; 2006 141B(3):261–268. [PubMed: 16526044]
26. Garbett K, Ebert PJ, Mitchell A, et al. Immune transcriptome alterations in the temporal cortex of subjects with autism. Neurobiol. Dis. Jun; 2008 30(3):303–311. [PubMed: 18378158]
27. Karlovich C, Duchateau-Nguyen G, Johnson A, et al. A longitudinal study of gene expression in healthy individuals. BMC Med Genomics. 2009; 2:33. [PubMed: 19500411]
28. McLoughlin K, Turteltaub K, Bankaitis-Davis D, et al. Limited dynamic range of immune response gene expression observed in healthy blood donors using RT-PCR. Mol Med. Jul-Aug; 2006 12(7–8):185–195. [PubMed: 17088951]
29. Campbell C, Vernon SD, Karem KL, Nisenbaum R, Unger ER. Assessment of normal variability in peripheral blood gene expression. Dis Markers. 2002; 18(4):201–206. [PubMed: 12590174]
30. Eady JJ, Wortley GM, Wormstone YM, et al. Variation in gene expression profiles of peripheral blood mononuclear cells from healthy volunteers. Physiol Genomics. Aug 11; 2005 22(3):402–411. [PubMed: 16014386]
31. Radich JP, Mao M, Stepaniants S, et al. Individual-specific variation of gene expression in peripheral blood leukocytes. Genomics. Jun; 2004 83(6):980–988. [PubMed: 15177552]
32. Pierce K, Glatt SJ, Liptak GS, McIntyre LL. The power and promise of identifying autism early: insights from the search for clinical and biological markers. Annals of Clinical Psychiatry. Jul-Sep;2009 21(3):132–147. [PubMed: 19758535]

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Table 1

Demographic and Clinical Characteristics of the Six Diagnostic Groups

	Diagnostic Group					
	AD	**PDD**	**DD**	**LD**	**TD**	**TIE**
Sample size, *n*	37	23	17	34	27	41
Sex, male:female	29:8	20:3	13:4	26:8	14:13	26:15
Age, range in months	13–46	12–41	13–43	12–32	12–45	13–38
Age mean months (s.d.)	26.9 (9.2)	28.7 (8.1)	25.0 (8.5)	17.4 (4.6)[a]	22.0 (9.0)	16.8 (5.8)[a]
Mullen Early Learning Composite Score, mean (s.d.)	69.2 (19.8)[b]	77.4 (14.8)[b]	63.6 (12.6)	94.3 (12.8)	116.6 (13.3)	105.4 (14.0)
Vineland Adaptive Behavior Scales Standard Score, mean (s.d.)	79.9 (12.2)[c]	82.6 (9.8)[c]	76.8 (15.2)	93.8 (7.7)	105.5 (9.7)	101.9 (9.5)
ADOS-T Social Affect Score, mean (s.d.)	14.8 (4.9)[d]	10.7 (3.2)[d]	6.8 (4.7)	4.0 (3.5)	1.8 (1.5)	2.5 (1.8)
ADOS-T Restricted/Repetitive Score, mean (s.d.)	3.1 (1.6)[e]	2.2 (1.5)[e]	1.4 (1.6)	0.9 (1.0)	0.6 (1.1)	0.5 (1.4)
ADOS-T Total Score, mean (s.d.)	17.9 (5.8)[f]	13.0 (3.2)	8.2 (5.4)	4.9 (3.8)	2.3 (1.8)	3.0 (2.5)

Abbreviations: AD = autistic disorder; ADOS-T = Autism Diagnostic Observation Schedule–Toddler; DD = developmental delay; LD = language delay; PDD = pervasive developmental disorder not otherwise specified; TD = typically developing; TIE = type-I error.

[a]Age: significantly different ($p<0.05$) from AD, DD, PDD, and TD groups

[b]Mullen: significantly different ($p<0.05$) from TD and TIE groups

[c]Vineland: significantly different ($p<0.05$) from LD, TD, and TIE groups

[d]ADOS-T Social Affect: significantly different ($p<0.05$) from all other diagnostic groups, including each other

[e]ADOS-T Restricted/Repetitive: significantly different ($p<0.05$) from all other diagnostic groups, but not each other

[f]ADOS-T Total: significantly different ($p<0.05$) from all other diagnostic groups, including each other

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 2

Probes Significantly Dysregulated ($p<0.05$, Fold-Change |1.2|) in the autism spectrum disorder (ASD) Discovery Sample Compared to the unaffected control group (CNT) Discovery Sample

Probe ID[a]	Gene Symbol	Gene Product	Log2 Fold-Change[b]	*p*
ILMN_2261600	*FCGR1A*	Fc fragment of IgG, high affinity Ia, receptor (CD64)	−1.69	1.29E-02
ILMN_1799848	*ANKRD22*	Ankyrin repeat domain 22	−1.65	2.63E-02
ILMN_2114568	*GBP5*	Guanylate binding protein 5	−1.65	3.89E-03
ILMN_2132599	*ANKRD22*	Ankyrin repeat domain 22	−1.64	2.77E-02
ILMN_1743145	*L-RAP*	*Homo sapiens* l-rap mRNA for leukocyte-derived arginine aminopeptidase long form variant, complete cds.	−1.62	3.91E-02
ILMN_2391051	*FCGR1B*	Fc fragment of IgG, high affinity Ib, receptor (CD64)	−1.62	8.29E-03
ILMN_2148785	*GBP1*	Guanylate binding protein 1, interferon-inducible, 67kDa	−1.60	1.04E-02
ILMN_2176063	*FCGR1A*	Fc fragment of IgG, high affinity Ia, receptor (CD64)	−1.59	2.31E-02
ILMN_1687306	*LGALS2*	Lectin, galactoside-binding, soluble, 2	−1.52	6.16E-03
ILMN_1694548	*ANXA3*	Annexin A3	−1.51	2.80E-03
ILMN_1786303	*LILRA3*	Leukocyte immunoglobulin-like receptor, subfamily A (without TM domain), member 3	−1.51	3.83E-02
ILMN_1701114	*GBP1*	Guanylate binding protein 1, interferon-inducible, 67kDa	−1.49	4.02E-02
ILMN_1708934	*ADM*	Adrenomedullin	−1.47	5.90E-03
ILMN_1745242	*PLSCR1*	Phospholipid scramblase 1	−1.47	3.18E-02
ILMN_1753111	*NAMPT*	Nicotinamide phosphoribosyltransferase	−1.46	4.22E-02
ILMN_1666078	*HLA-H*	Major histocompatibility complex, class I, H (pseudogene)	−1.41	2.66E-02
ILMN_1720771	*STX11*	Syntaxin 11	−1.41	2.77E-03
ILMN_1782487	*GBP1*	Guanylate binding protein 1, interferon-inducible, 67kDa	−1.40	3.14E-02
ILMN_1795336	*PTER*	Phosphotriesterase related	−1.40	1.33E-03
ILMN_1689734	*IL1RN*	Interleukin 1 receptor antagonist	−1.40	4.15E-02
ILMN_1771385	*BC050625*	Hs.409925	−1.39	2.35E-02
ILMN_2050911	*SLC22A4*	Solute carrier family 22 (organic cation/ergothioneine transporter), member 4	−1.39	4.21E-03
ILMN_1700067	*BTN3A2*	Butyrophilin, subfamily 3, member A2	−1.39	1.11E-02
ILMN_1748915	*S100A12*	S100 calcium binding protein A12	−1.38	2.58E-02
ILMN_1755843	*SLC26A8*	Solute carrier family 26, member 8	−1.37	7.40E-03
ILMN_1756806	*DQ893795*	Hs.719112	−1.35	1.92E-02
ILMN_1653871	*NAMPT*	Nicotinamide phosphoribosyltransferase	−1.35	1.85E-02
ILMN_1652003	*GNG10*	Guanine nucleotide binding protein (G protein), gamma 10	−1.35	4.91E-03
ILMN_1757074	*GNG10*	Guanine nucleotide binding protein (G protein), gamma 10	−1.35	9.47E-03
ILMN_1721116	*USP10*	Ubiquitin specific peptidase 10	−1.34	3.20E-03
ILMN_1731224	*PARP9*	Poly (ADP-ribose) polymerase family, member 9	−1.34	2.60E-02
ILMN_1679826	*CST7*	Cystatin F (leukocystatin)	−1.33	3.12E-02
ILMN_2335813	*GCH1*	GTP cyclohydrolase 1	−1.33	3.73E-02
ILMN_1676528	*BTN3A2*	Butyrophilin, subfamily 3, member A2	−1.33	3.01E-02
ILMN_1707979	*CARD17*	Caspase recruitment domain family, member 17	−1.33	3.26E-02
ILMN_1851599	*BX110640*	Unknown	−1.33	1.83E-02

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Probe ID[a]	Gene Symbol	Gene Product	Log2 Fold-Change[b]	*p*
ILMN_1696419	*STOM*	Stomatin	−1.33	2.07E-02
ILMN_2224486	*C3orf14*	Chromosome 3 open reading frame 14	−1.33	2.65E-03
ILMN_2053527	*PARP9*	Poly (ADP-ribose) polymerase family, member 9	−1.32	3.68E-02
ILMN_2357419	*LILRA5*	Leukocyte immunoglobulin-like receptor, subfamily A (with TM domain), member 5	−1.31	1.92E-02
ILMN_2294762	*AMY1A*	Amylase, alpha 1A (salivary)	−1.31	8.77E-03
ILMN_1685057	*SLC22A4*	Solute carrier family 22 (organic cation/ergothioneine transporter), member 4	−1.31	9.80E-03
ILMN_2256050	*SERPINA1*	Serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 1	−1.31	3.21E-02
ILMN_1709233	*F5*	Coagulation factor V (proaccelerin, labile factor)	−1.31	8.64E-04
ILMN_2126706	*LMNB1*	Lamin B1	−1.31	5.47E-03
ILMN_1810289	*MYOF*	Myoferlin	−1.30	4.60E-02
ILMN_1668865	*SLC2A14*	Solute carrier family 2 (facilitated glucose transporter), member 14	−1.30	3.54E-04
ILMN_1666742	*C9orf72*	Chromosome 9 open reading frame 72	−1.30	1.99E-02
ILMN_1767809	*GNG10*	Guanine nucleotide binding protein (G protein), gamma 10	−1.29	4.62E-03
ILMN_1726545	*LILRA5*	Leukocyte immunoglobulin-like receptor, subfamily A (with TM domain), member 5	−1.29	2.12E-02
ILMN_1684585	*ACSL1*	Acyl-CoA synthetase long-chain family member 1	−1.29	1.28E-02
ILMN_2393296	*DQ894773*	Hs.1466	−1.28	4.47E-02
ILMN_1725471	*DQ894773*	Hs.1466	−1.28	3.62E-02
ILMN_2109197	*EPB41L3*	Erythrocyte membrane protein band 4.1-like 3	−1.27	3.26E-02
ILMN_1726591	*CARD16*	Caspase recruitment domain family, member 16	−1.27	4.53E-02
ILMN_2219712	*HMGB2*	High-mobility group box 2	−1.26	1.55E-02
ILMN_1806015	*SLC2A3*	Solute carrier family 2 (facilitated glucose transporter), member 3	−1.26	5.79E-04
ILMN_1659047	*HIST2H2AA4*	Histone cluster 2, H2aa4	−1.26	3.13E-02
ILMN_1798270	*C11orf75*	Chromosome 11 open reading frame 75	−1.26	3.07E-02
ILMN_2232478	*APOBEC3G*	Apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3G	−1.26	2.62E-02
ILMN_1741727	*QPCT*	Glutaminyl-peptide cyclotransferase	−1.26	8.82E-03
ILMN_1772036	*STEAP4*	STEAP family member 4	−1.26	3.66E-02
ILMN_1710937	*IFI16*	Interferon, gamma-inducible protein 16	−1.25	3.15E-02
ILMN_1845037	*TRIM69*	Tripartite motif-containing 69	−1.25	3.98E-02
ILMN_1741881	*C9orf72*	Chromosome 9 open reading frame 72	−1.25	6.62E-03
ILMN_1795715	*DPYD*	Dihydropyrimidine dehydrogenase	−1.25	1.24E-02
ILMN_1737964	*HIATL1*	Hippocampus abundant transcript-like 1	−1.25	1.30E-02
ILMN_1680109	*COL4A3BP*	Collagen, type IV, alpha 3 (Goodpasture antigen) binding protein	−1.25	3.03E-02
ILMN_2412172	*APOBEC3F*	Apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3F	−1.25	2.16E-02
ILMN_1780546	*OSM*	Oncostatin M	−1.25	1.63E-02
ILMN_1779486	*FAM126B*	Family with sequence similarity 126, member B	−1.25	3.36E-02
ILMN_1674574	*VNN1*	Vanin 1	−1.24	3.86E-02
ILMN_2186061	*IPFK-2*	*Homo sapiens* inducible 6-phosphofructo-2-kinase/fructose 2,6-bisphosphatase (IPFK-2) mRNA, complete cds.	−1.23	1.66E-02
ILMN_1672124	*C4orf18*	Chromosome 4 open reading frame 18	−1.23	1.91E-02

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Probe ID[a]	Gene Symbol	Gene Product	Log2 Fold-Change[b]	p
ILMN_1788002	*MAPK14*	Mitogen-activated protein kinase 14	−1.23	6.14E-03
ILMN_1652680	*NCF1B*	Hs.647047	−1.23	4.82E-02
ILMN_1705663	*DMXL2*	Dmx-like 2	−1.23	4.59E-02
ILMN_1804735	*CBS*	Cystathionine-beta-synthase	−1.22	4.38E-02
ILMN_1674789	*Unknown*	Unknown	−1.22	4.52E-02
ILMN_1701914	*CD274*	CD274 molecule	−1.22	3.25E-02
ILMN_1713058	*PSTPIP2*	Proline-serine-threonine phosphatase interacting protein 2	−1.22	3.08E-02
ILMN_2061318	*TAF13*	TAF13 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 18kDa	−1.22	2.48E-02
ILMN_1683250	*LOC440731*	Similar to hCG1817424	−1.22	4.22E-02
ILMN_1659255	*RP2*	Retinitis pigmentosa 2 (X-linked recessive)	−1.22	4.25E-02
ILMN_2061327	*TAF13*	TAF13 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 18kDa	−1.22	1.90E-02
ILMN_1751079	*TAP1*	Transporter 1, ATP-binding cassette, sub-family B (MDR/TAP)	−1.22	2.42E-02
ILMN_1687201	*APOL6*	Apolipoprotein L, 6	−1.22	4.05E-02
ILMN_1667711	*DQ895783*	Unknown	−1.22	4.45E-02
ILMN_1801767	*ABHD3*	Abhydrolase domain containing 3	−1.22	3.78E-02
ILMN_1776649	*LRRK2*	Leucine-rich repeat kinase 2	−1.22	4.93E-02
ILMN_1781700	*IL18R1*	Interleukin 18 receptor 1	−1.21	1.31E-02
ILMN_2078264	*C3orf38*	Chromosome 3 open reading frame 38	−1.21	3.14E-02
ILMN_1654268	*HMGB2*	High-mobility group box 2	−1.21	1.93E-02
ILMN_2319077	*FAS*	Fas (TNF receptor superfamily, member 6)	−1.21	3.08E-02
ILMN_1761941	*C4orf18*	Chromosome 4 open reading frame 18	−1.21	2.23E-02
ILMN_1778059	*CASP4*	Caspase 4, apoptosis-related cysteine peptidase	−1.21	4.30E-02
ILMN_2226015	*LRRK2*	Leucine-rich repeat kinase 2	−1.21	3.97E-02
ILMN_1729115	*UBE2S*	Ubiquitin-conjugating enzyme E2S	−1.21	3.42E-02
ILMN_2230862	*GYG1*	Glycogenin 1	−1.21	1.06E-02
ILMN_2374340	*PLAUR*	Plasminogen activator, urokinase receptor	−1.21	4.33E-02
ILMN_1678766	*DYNLT1*	Dynein, light chain, Tctex-type 1	−1.21	4.65E-02
ILMN_1756953	*GBP6*	Guanylate binding protein family, member 6	−1.21	4.23E-02
ILMN_1792682	*MCTP2*	Multiple C2 domains, transmembrane 2	−1.21	1.96E-02
ILMN_1707286	*PLBD1*	Phospholipase B domain containing 1	−1.20	2.52E-02
ILMN_1736327	*CDC42EP3*	CDC42 effector protein (Rho GTPase binding) 3	−1.20	2.54E-02
ILMN_1716704	*NLRC5*	NLR family, CARD domain containing 5	−1.20	3.87E-03
ILMN_1806651	*PARP8*	Poly (ADP-ribose) polymerase family, member 8	−1.20	2.83E-02
ILMN_2189037	*WDR52*	WD repeat domain 52	−1.20	4.38E-02
ILMN_1691104	*PGAM4*	Phosphoglycerate mutase family member 4	−1.20	1.23E-02
ILMN_1685824	*B4GALT5*	UDP-Gal:betaGlcNAc beta 1,4- galactosyltransferase, polypeptide 5	−1.20	3.47E-02
ILMN_1745686	*MFHAS1*	Malignant fibrous histiocytoma amplified sequence 1	1.20	2.24E-03
ILMN_2320330	*MAL*	Mal, T-cell differentiation protein	1.20	3.54E-02
ILMN_2074860	*KRT4*	Keratin 4	1.21	1.54E-02
ILMN_1795639	*MGMT*	O-6-methylguanine-DNA methyltransferase	1.21	4.50E-03

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Probe ID[a]	Gene Symbol	Gene Product	Log2 Fold-Change[b]	*p*
ILMN_1713731	*ALOX12*	Arachidonate 12-lipoxygenase	1.21	3.71E-02
ILMN_1658624	*UBXN6*	UBX domain protein 6	1.21	4.99E-03
ILMN_1727183	*ZNF763*	Zinc finger protein 763	1.21	1.81E-02
ILMN_1772876	*ZNF395*	Zinc finger protein 395	1.21	4.67E-03
ILMN_1728107	*GNG7*	Guanine nucleotide binding protein (G protein), gamma 7	1.22	3.06E-03
ILMN_1658957	*MGC13005*	Hypothetical LOC84771	1.22	7.00E-03
ILMN_2402131	*TAF15*	TAF15 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 68kDa	1.23	2.83E-02
ILMN_1662587	*PNPLA7*	Patatin-like phospholipase domain containing 7	1.23	1.48E-02
ILMN_1662451	*FCER2*	Fc fragment of IgE, low affinity II, receptor for (CD23)	1.23	1.95E-02
ILMN_1676042	*Unknown*	Unknown	1.23	4.56E-02
ILMN_2337928	*CXCR5*	Chemokine (C-X-C motif) receptor 5	1.23	1.01E-02
ILMN_1753782	*ZNF266*	Zinc finger protein 266	1.23	1.24E-02
ILMN_2409384	*SIGLEC7*	Sialic acid binding Ig-like lectin 7	1.23	1.45E-02
ILMN_1762764	*SH3BGRL2*	SH3 domain binding glutamic acid-rich protein like 2	1.23	3.96E-02
ILMN_1666089	*OLIG1*	Oligodendrocyte transcription factor 1	1.24	3.21E-02
ILMN_2374352	*DBNDD1*	Dysbindin (dystrobrevin binding protein 1) domain containing 1	1.24	2.44E-02
ILMN_1659075	*HLA-DOA*	Major histocompatibility complex, class II, DO alpha	1.24	1.75E-02
ILMN_1773567	*LAMA5*	Laminin, alpha 5	1.24	4.66E-04
ILMN_2307025	*RBM12*	RNA binding motif protein 12	1.25	1.43E-02
ILMN_1735712	*KRT1*	Keratin 1	1.26	2.42E-02
ILMN_1657996	*AK098012*	Uncharacterized protein ENSP00000383353.	1.27	4.34E-03
ILMN_2143314	*SPIB*	Spi-B transcription factor (Spi-1/PU.1 related)	1.28	2.23E-02
ILMN_1695812	*KRT72*	Keratin 72	1.29	6.15E-03
ILMN_1685122	*COL9A2*	Collagen, type IX, alpha 2	1.29	8.81E-04
ILMN_1743911	*SLC25A39*	Solute carrier family 25, member 39	1.30	6.34E-03
ILMN_1752755	*VWF*	Von Willebrand factor	1.31	5.65E-03
ILMN_1703326	*GPR44*	G protein-coupled receptor 44	1.31	2.92E-02
ILMN_1814397	*EPB42*	Erythrocyte membrane protein band 4.2	1.32	3.25E-03
ILMN_1678707	*TAF15*	TAF15 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 68kDa	1.32	3.94E-02
ILMN_1763322	*CCR3*	Chemokine (C-C motif) receptor 3	1.32	2.97E-02
ILMN_1804350	*LOC644852*	Hypothetical LOC644852	1.33	2.42E-02
ILMN_1792323	*HDC*	Histidine decarboxylase	1.34	3.46E-02
ILMN_1751814	*BC032353*	*Homo sapiens* cDNA FLJ36366 fis, clone THYMU2007824.	1.35	2.06E-02
ILMN_1654875	*CLC*	Charcot-Leyden crystal protein	1.37	2.10E-02
ILMN_1796216	*VASH1*	Vasohibin 1	1.39	2.22E-04
ILMN_1705605	*BC032353*	*Homo sapiens* cDNA FLJ36366 fis, clone THYMU2007824.	1.39	4.81E-03
ILMN_1696512	*ERAF*	Erythroid associated factor	1.43	1.34E-02
ILMN_1815527	*HBD*	Hemoglobin, delta	1.51	8.33E-03
ILMN_1796678	*HBG1*	Hemoglobin, gamma A	2.04	9.24E-03
ILMN_2084825	*HBG2*	Hemoglobin, gamma G	2.07	7.00E-03

Note:

[a]The 48 probe IDs in **bold** font are those that comprised the optimal support vector machine (SVM) classifier of ASD.

[b]Rows are sorted by ascending fold-change in the ASD discovery sample compared to the CNT discovery sample, with largest negative fold-change at the top and largest positive fold-change at the bottom of the Table.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript