A meta-analysis of two high-risk prospective cohort studies reveals autism-specific transcriptional changes to chromatin, autoimmune, and environmental response genes in umbilical cord blood

Charles E. Mordaunt[1]* cemordaunt@ucdavis.edu

Bo Y. Park[2]* bopark@fullerton.edu

Kelly M. Bakulski[3] bakulski@umich.edu

Jason I. Feinberg[4] jfeinbe2@jhu.edu

Lisa A. Croen[5] Lisa.A.Croen@kp.org

Christine Ladd-Acosta[4] claddaco@jhsph.edu

Craig J. Newschaffer[7] cjn32@drexel.edu

Heather E. Volk[4] hvolk1@jhu.edu

Sally Ozonoff[7] sozonoff@ucdavis.edu

Irva Hertz-Picciotto[8] iher@ucdavis.edu

Janine M. LaSalle[1^+] jmlasalle@ucdavis.edu

Rebecca J. Schmidt[8^] rjschmidt@ucdavis.edu

M. Daniele Fallin[4^] dfallin@jhu.edu

*** First authors contributed equally to this study**

^ Last authors contributed equally to this study

+ Corresponding author

1. Department of Medical Microbiology and Immunology, Genome Center, and MIND Institute, University of California, Davis, CA, USA
2. Department of Public Health, California State University, Fullerton, CA, USA
3. Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI, USA
4. Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA

5. Division of Research and Autism Research Program, Kaiser Permanente Northern California, Oakland, CA, USA
6. Dornsife School of Public Health and A.J. Drexel Autism Institute, Drexel University, Philadelphia, PA, USA
7. Psychiatry and Behavioral Sciences and MIND Institute, University of California, Davis, USA
8. Department of Public Health Sciences and MIND Institute, University of California, Davis, USA

Abstract

Background: Autism spectrum disorder (ASD) is a neurodevelopmental disorder that affects more than 1% of children in the United States. ASD risk is thought to arise from a combination of genetic and environmental factors, with the perinatal period as a critical window. Understanding early transcriptional changes in ASD would assist in clarifying disease pathogenesis and identifying biomarkers and treatments. However, little is known about umbilical cord blood gene expression profiles in babies later diagnosed with ASD compared to non-typically developing (Non-TD) or neurotypical children.

Methods: Genome-wide transcript levels were measured by Affymetrix Human Gene 2.0 array in RNA from umbilical cord blood samples from both the Markers of Autism Risk in Babies--Learning Early Signs (MARBLES) and the Early Autism Risk Longitudinal Investigation (EARLI) high-risk pregnancy cohorts that enroll younger siblings of a child previously diagnosed with ASD. An algorithm-based diagnosis from 36 month assessments categorized the younger sibling as either ASD, typically developing (TD), or not ASD but non-typically developing (Non-TD). 59 ASD, 92 Non-TD, and 120 TD subjects were included and differences were identified in ASD versus TD subjects, with Non-TD versus TD as a specificity control. Meta-analysis was used to combine the results from both studies. Functional enrichments of differentially-expressed genes were examined across diagnostic groups.

Results: While cord blood gene expression differences comparing either ASD or Non-TD to TD did not reach genome-wide significance when adjusting for multiple comparisons, 172 genes were nominally differentially-expressed between ASD and TD cord blood (log_2(fold change) > 0.1, $p < 0.01$). These genes were significantly enriched for toxic substance response and xenobiotic metabolism functions, and gene sets involved in chromatin regulation and systemic lupus erythematosus were significantly upregulated (FDR $q < 0.05$). In contrast, 66 genes were differentially-expressed between Non-TD and TD cord blood, including only 8 genes that were also differentially-expressed in ASD.

Conclusions: This is the first study to identify perinatal gene expression differences in umbilical cord blood specific to ASD. The results of this meta-analysis across two prospective ASD cohorts support involvement of environmental, immune, and epigenetic mechanisms in ASD etiology.

Keywords

autism spectrum disorder, neurodevelopment, umbilical cord blood, gene expression, meta-analysis, prospective study, microarray, chromatin, environment, autoimmune, perinatal

Background

Autism spectrum disorder (ASD) is a neurodevelopmental disorder characterized by impaired social interaction and restricted and repetitive behaviors. Heritability of ASD risk has been well established with twin and family studies and estimated at 52% [1-3]. While rare variants with large effects explain a relatively small proportion of all ASD cases, heritable common variants with individually minor effects contribute substantially to ASD risk [4]. Accumulating lines of evidence suggest that ASD arises from complex interactions between heterogeneous genetic and environmental risk factors. Gene expression levels are influenced by both genetic and environmental factors and determine the functional responses of cells and tissues. Postmortem

brain gene expression studies have guided understanding of ASD pathophysiology and show evidence of changes in gene co-expression and enrichment in immune response and neuronal activity functions [5, 6]. Peripheral blood gene expression studies in children and adults using whole blood and in specific cell types (natural killer (NK) cell and lymphocytes) observed enrichment of immune and inflammatory processes in differential gene expression associated with ASD [7, 8]. Recent efforts have been focused on identifying how genetic risk factors converge into one or more unifying pathways and pathophysiological mechanisms [9, 10]. Yet, the majority of this work to date relies on post-mortem or post-symptom timing of sample collection, rather than prospective assessment of gene expression.

Converging evidence suggests that most of the changes in the brain associated with ASD are initiated during prenatal brain development [11, 12], but the complete nature of these changes remain unknown. Umbilical cord blood captures fetal blood as well as the exchanges across the feto-placental unit and provides a distinct insight into prenatal development. A unique cell mixture is represented in umbilical cord blood, including hematopoietic stem cells, B cells, NK cells, T cells, monocytes, granulocytes and nucleated red blood cells [13]. Cord blood gene expression would reflect the immune response as well as endocrine and cellular communication essential for fetal development near the time of birth.

While several studies have previously examined child blood gene expression differences in ASD [8, 14-20], this is the first study to take advantage of cord blood samples collected from two prospective studies (Markers of Autism Risk in Babies--Learning Early Signs (MARBLES) and the Early Autism Risk Longitudinal Investigation (EARLI)) in order to assess the perinatal transcriptional changes that precede ASD diagnosis in high-risk children [21, 22]. The subjects in this study are all siblings of children with ASD and thus have a 13-fold increased risk for ASD compared to the general population [23]. They are also at a higher risk for non-typical

neurodevelopment, including deficits in attention and behavior. We measured cord blood gene expression levels using the Affymetrix Human Gene 2.0 array and compared the gene-level differential expression and gene set enrichment across ASD, non-typically developing (Non-TD) and neurotypical children (Fig. S1). Study-level results were then combined in a meta-analysis to investigate cord blood transcriptional dysregulation in ASD.

Methods

Sample population and biosample collection

MARBLES

The MARBLES study recruits Northern California mothers from lists of children receiving services through the California Department of Developmental Services who have a child with confirmed ASD and are planning a pregnancy or are pregnant with another child. Inclusion criteria for the study were: 1) mother or father has one or more biological child(ren) with ASD; 2) mother is 18 years or older; 3) mother is pregnant; 4) mother speaks, reads, and understands English sufficiently to complete the protocol and the younger sibling will be taught to speak English; 5) mother lives within 2.5 hours of the Davis/Sacramento region at time of enrollment. As described in more detail elsewhere [21], demographic, diet, lifestyle, environmental, and medical information were prospectively collected through telephone-assisted interviews and mailed questionnaires throughout pregnancy and the postnatal period. Mothers were provided with sampling kits for cord blood collection prior to delivery. MARBLES research staff made arrangements with obstetricians/midwives and birth hospital labor and delivery staff to assure proper sample collection and temporary storage. Infants received standardized neurodevelopmental assessments beginning at 6 months, as described below, and concluding at 3 years of age. For this study, all children actively enrolled by March 1, 2017 (n = 347) with umbilical cord blood collected in a PAXgene Blood RNA tube (n = 262, 76%) were included.

EARLI

The EARLI study is a high-risk pregnancy cohort that recruited and followed pregnant mothers who had an older child diagnosed with ASD through pregnancy, birth, and the first three years of life. EARLI families were recruited at four EARLI Network sites (Drexel/Children's Hospital of Philadelphia; Johns Hopkins/Kennedy Krieger Institute; University of California (UC) Davis; and Kaiser Permanente Northern California) in three distinct US regions (Southeast Pennsylvania, Northeast Maryland, and Northern California). In addition to having a biological child with ASD confirmed by EARLI study clinicians, to be eligible mothers also had to communicate in English or Spanish and, at recruitment, meet the following criteria: be 18 years or older; live within two hours of a study site; and be less than 29 weeks pregnant. The design of the EARLI study is described in more detail in Newschaffer et al. [22]. EARLI research staff made arrangements with obstetricians/midwives and birth hospital labor and delivery staff to ensure proper cord blood sample collection and temporary storage. The development of children born into the cohort was closely followed through age three years. For this study, 212 infants born into EARLI as a singleton birth and followed to one year of age were considered for inclusion. Of the 212 infants, 97 were excluded because they were either missing umbilical cord blood samples or outcome measures at 36 months, leaving a final sample of 115.

Diagnostic outcomes

In both studies, development was assessed by trained, reliable examiners. Diagnostic assessments at three years included the gold standard Autism Diagnostic Observation Schedule (ADOS) [24, 25], the Autism Diagnostic Interview-Revised (ADI-R) [26] conducted with parents, and the Mullen Scales of Early Learning (MSEL) [27], a test of cognitive, language, and motor development. Participants were classified into one of three outcome groups, ASD, typically developing (TD), and Non-TD, based on a previously published algorithm that uses ADOS and MSEL scores [28, 29]. Children with ASD outcomes had scores over the

ADOS cutoff and met DSM-5 criteria for ASD. The Non-TD group was defined as children with low MSEL scores (i.e., two or more MSEL subscales that are more than 1.5 standard deviations (SD) below average or at least one MSEL subscale that was more than 2 SD below average), elevated ADOS scores (i.e., within 3 points of the ASD cutoff), or both. Children with TD outcomes had all MSEL scores within 2 SD and no more than one MSEL subscale 1.5 SD below the normative mean and scores on the ADOS at least three or more points below the ASD cutoff.

RNA isolation and expression assessment

In both EARLI and MARBLES, umbilical cord blood was collected at the time of birth in PAXgene Blood RNA Tubes with the RNA stabilization reagent (BD Biosciences) and stored at -80°C. RNA isolation was performed with the PAXgene Blood RNA Kit (Qiagen) following the manufacturer's protocol. RNA from 236 (90%) of the 262 MARBLES PAXgene blood samples and all of the EARLI PAXgene blood samples met quality control standards (RIN ≥ 7.0 and concentration ≥ 35ng/uL) and volume requirements. Total RNA was converted to cDNA and *in vitro* transcribed to biotin-labeled cRNA, which was hybridized to Human Gene 2.0 Affymetrix microarray chips by the Johns Hopkins Sequencing and Microarray core. EARLI and MARBLES samples were measured separately and in multiple batches within each study. The manufacturer's protocol was followed for all washing, staining and scanning procedures. The raw fluorescence data (in Affymetrix CEL file format) with one perfect match and one mismatched probe in each set were analyzed using oligo package in R.

Data preprocessing

Within each study, signal distribution was first assessed in perfect-match probe intensity and Robust Multi-Chip Average (RMA) normalized data [30]. During the quality control step, we identified outliers using the arrayQualityMetrics and oligo R packages [31, 32]. Outliers were

excluded based on loading in principal component 1, the Kolmogorov-Smirnov test, median normalized unscaled standard error, and the sum of the distances to all other arrays. For the MARBLES study, 3 outlier samples were identified and excluded, and another 71 children had not yet received a diagnosis by April 12, 2018 so were excluded; 162 samples were normalized using RMA. For the EARLI study, 6 outliers were identified and excluded, then 109 samples were normalized using RMA. Probes were annotated at the transcript level using the pd.hugene.2.0.st R package [33], and those assigned to a gene (36,459 probes) were used in subsequent analyses.

Differential gene expression

We used surrogate variable analysis (SVA) to estimate and adjust for unmeasured environmental, demographic, cell type proportion, and technical factors that may have substantial effects on gene expression using the SVA R package [34]. 21 and 11 surrogate variables were detected in normalized expression data from the MARBLES and EARLI studies, respectively. Specific factors associated with surrogate variables in this study included sex and array batch. Differential expression was determined using the limma package in R with diagnosis and surrogate variables included in the linear model [35] (Fig. S2, S3). ASD versus TD and Non-TD versus TD differential expression results were extracted from a single model with three levels for diagnosis. Fold change and standard error from each study were input into the METAL command-line tool for meta-analysis [36]. Using the meta-analyzed data, differential probes were then identified as those with a nominal *p*-value less than 0.01 and an average absolute $\log_2$(fold change) greater than 0.1.

Gene overlap analysis

Gene overlap analysis by Fisher's exact test was performed using the GeneOverlap R package [37]. Gene symbols annotated to differentially-expressed probes were compared to autism-

related or blood cell-type associated gene lists [38] for overlap relative to all genes annotated to probes on the array. Genes with variation previously associated with autism were obtained from the Simons Foundation Autism Research Initiative (SFARI) Gene database and a recent genome-wide association study meta-analysis [39, 40], while genes with expression previously associated with autism were obtained from multiple previous reports [6, 8, 41, 42]. Significant overlaps were those with a false discovery rate (FDR) q-value < 0.05.

Overrepresentation enrichment analysis

Differential probes identified during meta-analysis were converted to Entrez gene IDs using the biomaRt R package [43]. Functional enrichment of only differential probes by hypergeometric test was relative to all probes on the array and was performed using the WebGestalt online tool with default parameters for the overrepresentation enrichment analysis method [44]. Enrichment databases included WebGestalt defaults and also a custom database of recently evolved genes obtained from [45]. WebGestalt default databases queried included Gene Ontology, KEGG, WikiPathways, Reactome, PANTHER, MSigDB, Human Phenotype Ontology, DisGeNET, OMIM, PharmGKB, and DrugBank. Significant enrichments were those with an FDR q-value < 0.05.

Gene set enrichment analysis (GSEA)

All probes included in the analysis were ranked using meta-analysis $\log_2$(fold change) and input into the WebGestalt online tool using default parameters for the GSEA method [44]. GSEA assesses whether genes in biologically-predefined sets occur toward the top or bottom of a ranked list of all examined genes more than expected by chance [46]. GSEA calculates an enrichment score normalized to the set size to estimate the extent of non-random distribution of the predefined gene set, and it then tests the significance of the enrichment with a permutation test. Enrichment databases included WebGestalt defaults (see above) and also a custom

database of blood cell-type associated genes [38]. Significant gene sets were called as those with an FDR *q*-value < 0.05.

Results

Study Sample Characteristics

MARBLES subjects in the final analysis included 77 TD (40 male, 37 female), 41 ASD (30 male, 11 female), and 44 Non-TD subjects (27 male, 17 female). Paternal age and gestational age were nominally associated with diagnostic group in MARBLES, with slightly increased paternal age and gestational age for the ASD subjects (paternal age $p = 0.02$, gestational age $p = 0.04$, Table 1). EARLI subjects in the final analysis included 43 TD (19 male, 24 female), 18 ASD (13 male, 5 female), and 48 Non-TD subjects (23 male, 25 female). Child race and ethnicity and home ownership were nominally associated with diagnostic group in EARLI (race and ethnicity $p = 0.02$, home ownership $p = 0.01$, Table 2). Specifically, the ASD group included a lower proportion of white subjects and a lower rate of home ownership. In the meta-analysis, which combined both the MARBLES and EARLI studies, gene expression was analyzed in 271 subjects, including 120 TD, 59 ASD, and 92 Non-TD subjects.

Table 1. Demographic characteristics of children and their parents in the MARBLES study, stratified by child diagnosis.

	Child 36 Month Diagnosis			
	ASD (*n* = 41)	**Non-TD (*n* = 44)**	**TD (*n* = 77)**	***p*-value[a]**
Child male gender, *n* (%)	30 (73.2)	27 (61.4)	40 (51.9)	0.08
Child race and ethnicity, *n* (%)				0.98
White (European or Middle Eastern)	19 (46.3)	18 (40.9)	38 (49.4)	
Black/African-American	2 (4.9)	2 (4.5)	1 (1.3)	
Asian	4 (9.8)	3 (6.8)	6 (7.8)	
Pacific Islander	0 (0)	0 (0)	1 (1.3)	
Hispanic White (1 parent)	5 (12.2)	5 (11.4)	11 (14.3)	
Hispanic White (2 parents)	3 (7.3)	6 (13.6)	8 (10.4)	
Hispanic Non-White (1 parent)	2 (4.9)	1 (2.3)	3 (3.9)	
Hispanic Non-White (2 parents)	2 (4.9)	2 (4.5)	3 (3.9)	
Multiracial	4 (9.8)	7 (15.9)	6 (7.8)	
Mother age (years), *mean* (*SD*)[b]	35.4 (5.1)	34.1 (4.3)	33.3 (4.9)	0.09
Father age (years), *mean* (*SD*)[c]	38.7 (5.3)	36.1 (4.8)	35.7 (6.2)	**0.02**
Mother pre-pregnancy BMI, *mean* (*SD*)[d]	27.9 (7.4)	29.3 (8.3)	26.1 (6.2)	0.05
Mother urinary cotinine (ng/mL), *mean* (*SD*)[e]	115.7 (324.8)	18.9 (89.1)	6.1 (33.9)	0.07
Mother bachelor's degree +, *n* (%)	18 (43.9)	19 (43.2)	40 (51.9)	0.57
Father bachelor's degree +, *n* (%)	21 (51.2)	19 (43.2)	39 (50.6)	0.69
Own home, *n* (%)	21 (51.2)	25 (56.8)	48 (62.3)	0.43
Gestational age (weeks), *mean* (*SD*)	39.4 (1.1)	39 (1.3)	38.7 (1.4)	**0.04**
Birth weight (kg), *mean* (*SD*)	3.5 (0.4)	3.5 (0.4)	3.4 (0.5)	0.52

Autism spectrum disorder (ASD), Non-typically developing (Non-TD), Typically developing (TD);

[a] *p*-values from Fisher's exact test for categorical variables and one-way ANOVA for continuous variables; [b]Frequency missing = 1 ASD; [c]Frequency missing = 2 ASD, 1 TD; [d]Frequency missing = 1 ASD, 1 TD; [e]Frequency missing = 18 ASD, 21 Non-TD, 42 TD;

Table 2. Demographic characteristics of children and their parents in the EARLI study, stratified by child diagnosis.

	Child 36 Month Diagnosis			
	ASD (*n* = 18)	**Non-TD (*n* = 48)**	**TD (*n* = 43)**	***p*-value[a]**
Child male gender, *n* (%)	13 (72.2)	23 (47.9)	19 (44.2)	0.13
Child race and ethnicity, *n* (%)[b]				**0.02**
White (European or Middle Eastern)	6 (35.3)	23 (48.9)	24 (55.8)	
Black/African-American	3 (17.6)	8 (17)	1 (2.3)	
Asian	2 (11.8)	5 (10.6)	0 (0)	
Hispanic White (1 parent)	0 (0)	1 (2.1)	2 (4.7)	
Hispanic White (2 parents)	2 (11.8)	2 (4.3)	1 (2.3)	
Hispanic Non-White (1 parent)	2 (11.8)	3 (6.4)	3 (7)	
Hispanic Non-White (2 parents)	0 (0)	4 (8.5)	5 (11.6)	
Multiracial	2 (11.8)	1 (2.1)	7 (16.3)	
Mother age (years), *mean* (*SD*)	34.4 (4)	33.8 (4.3)	34.7 (5)	0.60
Father age (years), *mean* (*SD*)	35.4 (6.8)	35.7 (5)	36.2 (6.4)	0.88
Mother pre-pregnancy BMI, *mean* (*SD*)[c]	29.4 (7.1)	27.4 (7)	28 (7.2)	0.60
Mother urinary cotinine (ng/mL), *mean* (*SD*)[d]	14.3 (55.7)	16.5 (107.5)	0.3 (0.3)	0.59
Mother bachelor's degree +, *n* (%)	13 (72.2)	44 (91.7)	39 (90.7)	0.10
Father bachelor's degree +, *n* (%)[e]	13 (72.2)	38 (79.1)	33 (76.7)	0.43
Own home, *n* (%)	8 (44.4)	31 (64.6)	36 (83.7)	**0.01**
Gestational age (weeks), *mean* (*SD*)	39.3 (1.4)	39.4 (1.4)	39.1 (1.5)	0.56
Birth weight (kg), *mean* (*SD*)[f]	3.5 (0.6)	3.5 (0.6)	3.5 (0.6)	1.00

Autism spectrum disorder (ASD), Non-typically developing (Non-TD), Typically developing (TD);

[a]*p*-values from Fisher's exact test for categorical variables and one-way ANOVA for continuous variables; [b]Frequency missing = 1 ASD, 1 Non-TD; [c]Frequency missing = 1 Non-TD; [d]Frequency missing = 2 ASD, 4 Non-TD, 3 TD; [e]Frequency missing = 3 Non-TD; [f]Frequency missing = 1 Non-TD;

ASD-Associated Differential Gene Expression in Cord Blood

We examined differential expression of single genes in association with ASD diagnosis status at 36 months. In the meta-analysis, no transcripts were differentially expressed at a conservative FDR *q*-value < 0.05. Under the thresholds of $\log_2$(fold change) > 0.1 and nominal *p*-value <

0.01, 172 transcripts were differentially expressed between ASD and TD cord blood (ASD *n* = 59, TD *n* = 120, Fig. 1a, Table S1). Among these, 87 were upregulated and 85 were downregulated. The differential transcript with the greatest fold change was *TUBB2A* ($\log_2$(fold change) = 0.35, Fig. 1b, Table 3). Additionally, the estimated fold changes for differentially-expressed genes were correlated between the two studies (Pearson's r = 0.80, p < 2.2E-16). Many of the differentially-expressed genes were noncoding or uncharacterized transcripts; however, the median expression of differentially-expressed genes was not lower than non-differentially-expressed genes on the array (MARBLES: differential = 4.70, non-differential = 4.64, p = 0.74; EARLI: differential = 4.34, non-differential = 4.19, p = 0.52; Fig. S4).

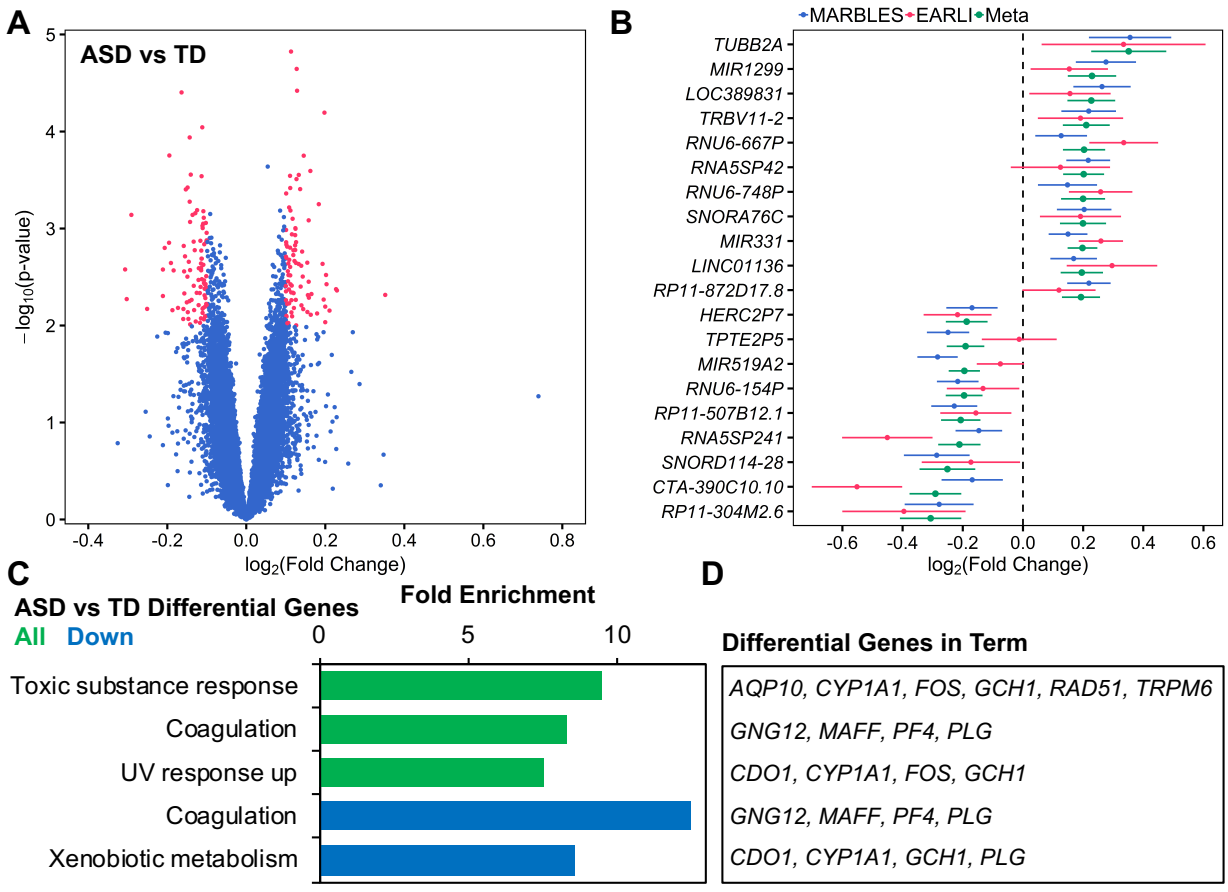

Figure 1. Identification and function of ASD-associated differentially-expressed genes in cord blood from two high-risk prospective studies. Gene expression in umbilical cord

blood samples from subjects with typical development (n = 120, 59 male/61 female) or those diagnosed with ASD at age 36 months (n = 59, 43 male/16 female) was assessed by expression microarray. SVA was performed to control for technical and biological variables including sex and array batch. Differential expression analysis was carried out separately by study and combined in a meta-analysis. (A) Identification of 172 differentially-expressed genes in meta-analysis (178 probes, $\log_2$(fold change) > 0.1, p < 0.01). (B) Fold change in gene expression for top 20 differentially-expressed genes sorted by meta-analysis $\log_2$(fold change) and plotted for individual studies and meta-analysis. (C) ASD-associated differentially-expressed genes were analyzed for functional enrichment with WebGestalt using the hypergeometric test and compared to all genes annotated to the array. Significantly-enriched ontology terms are shown (FDR q < 0.05). (D) ASD-associated differentially-expressed genes belonging to significantly-enriched ontology terms are listed.

Table 3. Top 20 ASD-associated differentially-expressed genes by log_2(fold change) in meta-analysis.

		MARBLES		EARLI		Meta-Analysis			
Probe ID	Gene Symbol	log_2(FC)	*p*-value	log_2(FC)	*p*-value	log_2(FC)	SE	*p*-value	*q*-value
17004333	*TUBB2A*	0.36	9.6E-03	0.33	2.1E-01	0.35	0.12	4.8E-03	0.75
17094496	*MIR1299*	0.28	6.1E-03	0.15	2.3E-01	0.23	0.08	4.3E-03	0.75
17065047	*LOC389831*	0.26	6.1E-03	0.16	2.4E-01	0.23	0.08	4.2E-03	0.74
17063807	*TRBV11-2*	0.22	1.7E-02	0.19	1.7E-01	0.21	0.08	7.0E-03	0.79
16812322	*RNU6-667P*	0.13	1.4E-01	0.34	4.1E-03	0.20	0.07	3.7E-03	0.73
16662296	*RNA5SP42*	0.22	3.1E-03	0.12	4.4E-01	0.20	0.07	3.0E-03	0.73
17079513	*RNU6-748P*	0.15	1.3E-01	0.26	1.5E-02	0.20	0.07	6.4E-03	0.78
16837063	*SNORA76C*	0.20	2.5E-02	0.19	1.5E-01	0.20	0.08	9.2E-03	0.83
16755265	*MIR331*	0.15	2.1E-02	0.26	7.0E-04	0.20	0.05	6.4E-05	0.47
16698226	*LINC01136*	0.17	3.0E-02	0.30	4.9E-02	0.20	0.07	5.4E-03	0.76
16738477	*RP11-872D17.8*	0.22	2.6E-03	0.12	3.2E-01	0.19	0.06	2.3E-03	0.72
16797869	*HERC2P7*	-0.17	4.7E-02	-0.22	5.4E-02	-0.19	0.07	7.0E-03	0.79
16778413	*TPTE2P5*	-0.25	4.9E-04	-0.01	9.2E-01	-0.19	0.06	2.3E-03	0.72
16865104	*MIR519A2*	-0.28	3.5E-05	-0.07	3.3E-01	-0.19	0.05	1.8E-04	0.70
17113511	*RNU6-154P*	-0.22	1.9E-03	-0.13	2.6E-01	-0.20	0.06	1.4E-03	0.72
16810171	*RP11-507B12.1*	-0.23	3.1E-03	-0.16	1.8E-01	-0.21	0.07	1.6E-03	0.72
17062251	*RNA5SP241*	-0.15	5.8E-02	-0.45	3.4E-03	-0.21	0.07	2.6E-03	0.72
16788697	*SNORD114-28*	-0.29	9.1E-03	-0.17	2.8E-01	-0.25	0.09	6.7E-03	0.78
16928422	*CTA-390C10.10*	-0.17	9.8E-02	-0.55	4.3E-04	-0.29	0.09	7.2E-04	0.72
16733787	*RP11-304M2.6*	-0.28	1.5E-02	-0.40	5.3E-02	-0.31	0.10	2.6E-03	0.72

Autism spectrum disorder (ASD), fold change (FC), standard error (SE);

Several differentially-expressed genes in cord blood overlapped with genes previously associated with ASD in genetic or gene expression studies, including *SLC7A3*, *VSIG4*, *MIR1226*, *SNORD104*, *OR2AG2*, and *DHX30*, although the overlap was not statistically significant ($q > 0.05$, Fig. S5, Table S2). Two known ASD genes in the SFARI Gene database, *SLC7A3* and *VSIG4*, were upregulated in cord blood from ASD subjects (*SLC7A3*: log_2(fold change) = 0.16, p = 2.6E-4; *VSIG4*: log_2(fold change) = 0.12, p = 2.2E-3) [39]. Additionally, *MIR1226*, previously associated with ASD in a large genome-wide association study meta-analysis, was downregulated in ASD subject cord blood (log_2(fold change) = -0.14, p = 8.6E-4) [40]. *SNORD104* was upregulated and *OR2AG2* was downregulated in cord blood, and both

were differentially expressed in the same direction in a previous ASD expression meta-analysis in blood (*SNORD104*: $\log_2$(fold change) = 0.20, p = 9.2E-3; *OR2AG2*: $\log_2$(fold change) = -0.14, p = 2.8E-4) [8]. *DHX30* was downregulated in cord blood and also in a previous autism expression study in post-mortem cortex ($\log_2$(fold change) = -0.14, p = 8.6E-4) [41]. *GFI1*, *GPR171*, *KIR2DL4, PTGIR*, and *TRPM6* were differentially expressed in ASD cord blood and are also differentially expressed in specific blood cell types, including natural killer cells and T cells, although a significant enrichment was not observed (q > 0.05, Fig. S6, Table S3) [38].

Overrepresentation enrichment analysis, which looks for overlap of only differentially-expressed genes with biologically-predefined gene lists, revealed that ASD differential transcripts were significantly enriched for functions in the response to toxic substances (fold enrichment = 9.5, q = 0.027) and ultraviolet radiation (fold enrichment = 7.6, q = 0.037, Fig. 1c, Table S4). Genes associated with both of these annotations included *CYP1A1, FOS*, and *GCH1*. Additionally, downregulated transcripts were enriched for functioning in blood coagulation (*GNG12*, *MAFF, PF4*, and *PLG,* fold enrichment = 12.5, q = 0.009) and xenobiotic metabolism (*CDO1*, *CYP1A1*, *GCH1*, and *PLG,* fold enrichment = 8.6, q = 0.019).

Using fold changes to rank all transcripts for gene set enrichment analysis (GSEA), we observed significant enrichment for upregulation of gene sets involved in chromatin regulation (q < 0.05, Fig. 2, Table S5). In other words, genes associated with chromatin regulation tended to be ranked toward the top of the distribution of fold change in ASD cord blood. Chromatin gene sets upregulated in ASD included DNA methylation (normalized enrichment score (NES) = 2.16, q = 0.009), condensation of prophase chromosomes (NES = 2.11, q = 0.007), nucleosome assembly (NES = 1.96, q = 0.026), histone deacetylase (HDAC)-mediated deacetylation (NES = 1.90, q = 0.040), and polycomb-repressive complex 2 (PRC2)-mediated methylation (NES = 1.89, q = 0.002). Additionally, the gene set for the autoimmune disease systemic lupus

erythematosus was significantly upregulated (NES = 2.13, q = 0.003). Most of the genes associated with these sets compose a cluster of histone genes located at the 6p22.2 locus, which was also enriched (NES = 2.15, q = 0.007). The above findings of differential expression across two prospective cohorts suggest transcriptional dysregulation in environmentally-responsive genes is present at birth in cord blood of high-risk subjects later diagnosed with ASD.

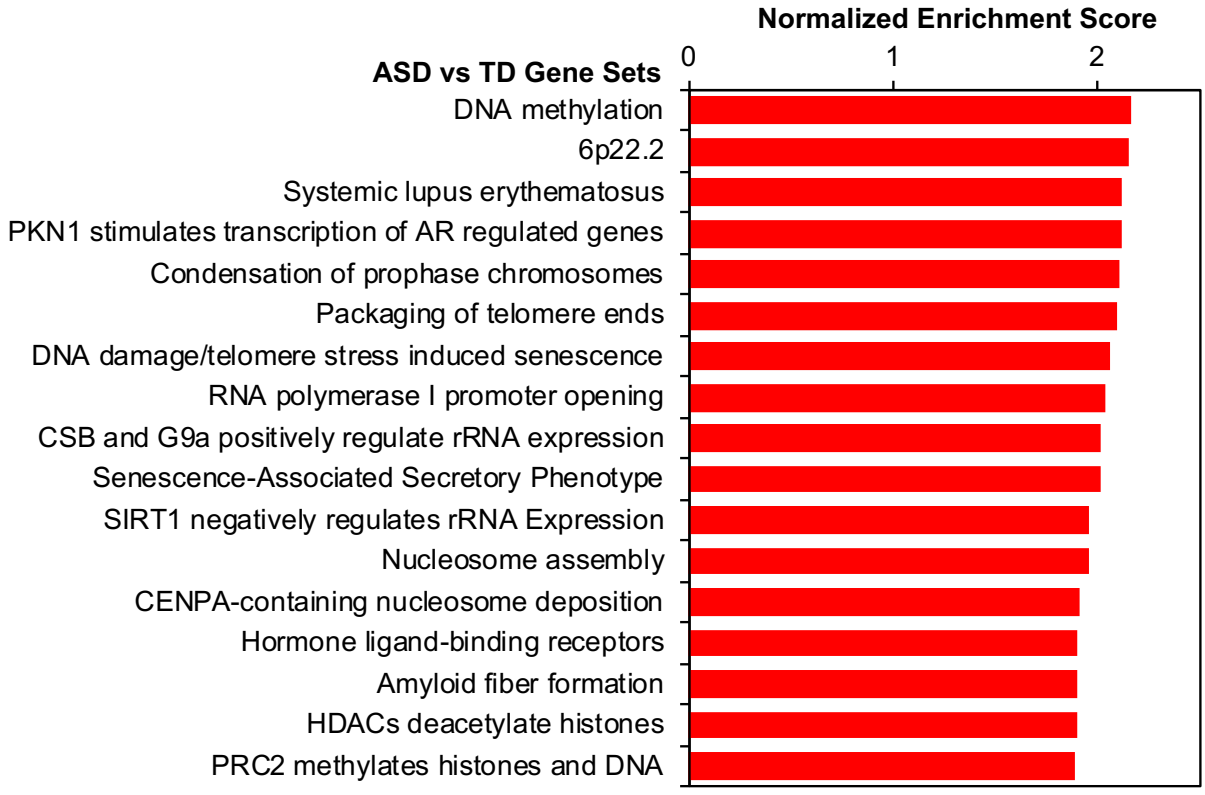

Figure 2. Chromatin and autoimmune gene sets are upregulated in cord blood from ASD subjects. Meta-analysis $\log_2$(fold change) for ASD versus TD gene expression was used to rank probes for gene set enrichment analysis (GSEA) with WebGestalt. GSEA assesses whether genes in biologically-defined sets occur toward the top or bottom of a ranked list more than expected by chance. Significantly-enriched gene sets are shown (FDR q < 0.05).

Non-TD-Associated Differential Gene Expression in Cord Blood

To assess the specificity of ASD-associated transcriptional changes in cord blood, we also examined differential expression between Non-TD and TD cord blood samples. Meta-analysis results showed no transcripts differentially expressed at a conservative FDR q-value < 0.05. Under the thresholds of $\log_2$(fold change) > 0.1 and nominal p-value < 0.01, 66 transcripts were differential, with 38 upregulated and 28 downregulated (Non-TD n = 92, TD n = 120, Fig. 3a, Table S6). The gene with the greatest fold change between Non-TD and TD subjects was *TAS2R46* ($\log_2$(fold change) = 0.37, Fig. 3b, Table 4). Further, the estimated fold changes of Non-TD-associated differentially-expressed genes were correlated between the individual studies (Pearson's r = 0.80, p = 3.9E-16). Although many of the differential transcripts have limited known functions in cord blood, median expression of these genes was not different from other genes on the array (MARBLES: differential = 4.48, non-differential = 4.64, p = 0.65; EARLI: differential = 4.15, non-differential = 4.20, p = 0.90; Fig. S7).

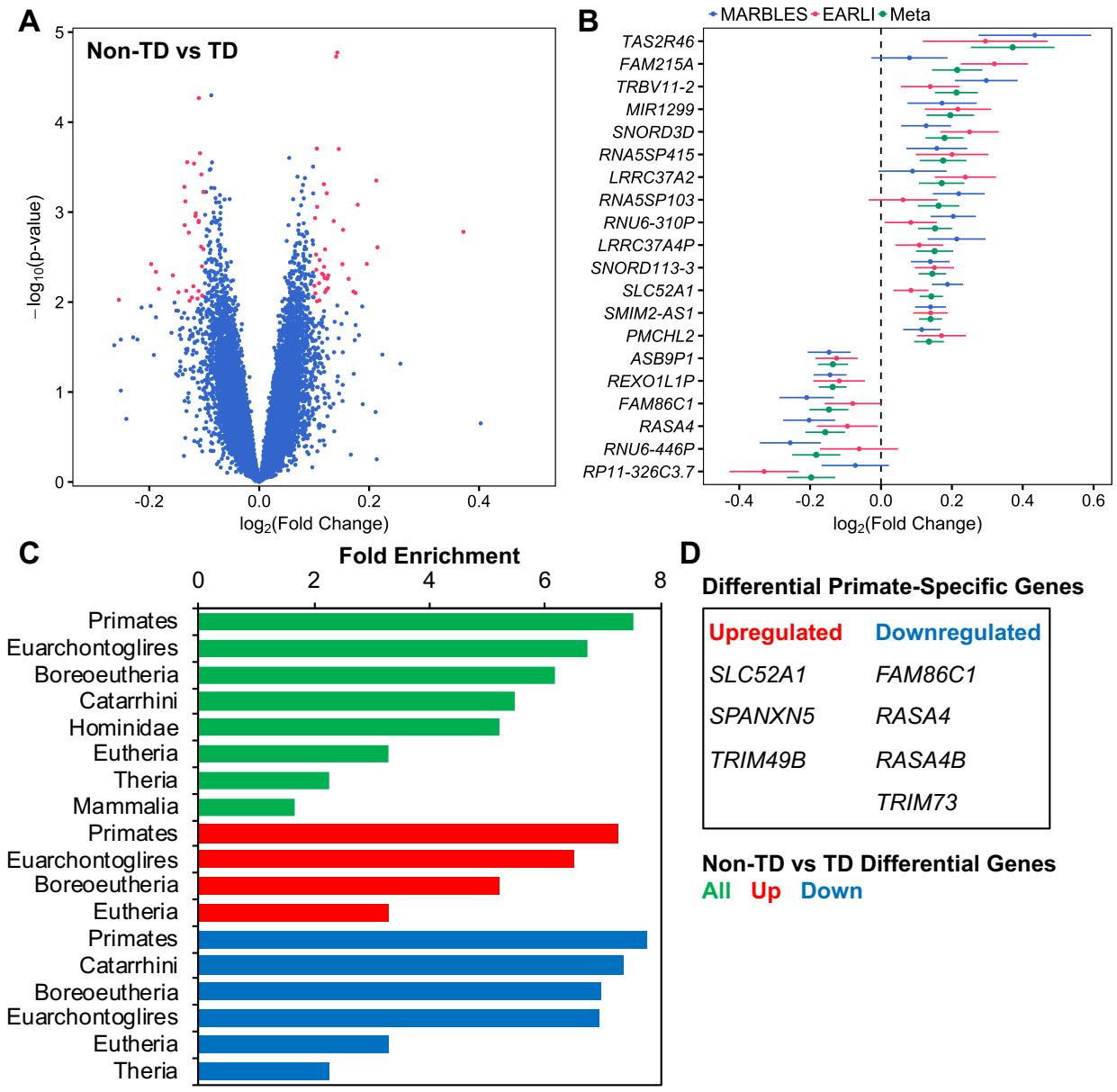

Figure 3. Identification and functional enrichment of Non-TD-associated differentially-expressed genes in cord blood. Gene expression in umbilical cord blood samples from subjects with typical development (TD, *n* = 120, 59 male/61 female) or those diagnosed with Non-TD at age 3 (*n* = 92, 50 male/42 female) was assessed by expression microarray. SVA was performed to control for technical and biological variables including sex and array batch. Differential expression analysis was carried out separately by study and combined in a meta-analysis. (A) Identification of 66 differentially-expressed genes in meta-analysis (66

probes, $\log_2$(fold change) > 0.1, $p < 0.01$). (B) Fold change in gene expression for top 20 differentially-expressed genes by meta-analysis $\log_2$(fold change) for individual studies and meta-analysis. (C) Non-TD-associated differentially-expressed genes were analyzed for functional enrichment in recently evolved genes with WebGestalt using the hypergeometric test. Significantly-enriched clades are shown (FDR $q < 0.05$). (D) Non-TD-associated differentially-expressed genes specific to primates are listed.

Table 4. Top 20 Non-TD-associated differentially-expressed genes by $\log_2$(fold change) in meta-analysis.

		MARBLES		EARLI		Meta-Analysis			
Probe ID	Gene Symbol	$\log_2$(FC)	*p*-value	$\log_2$(FC)	*p*-value	$\log_2$(FC)	SE	*p*-value	*q*-value
16761521	*TAS2R46*	0.43	6.7E-03	0.29	9.3E-02	0.37	0.12	1.7E-03	0.78
16846019	*FAM215A*	0.08	4.5E-01	0.32	9.5E-04	0.22	0.07	2.5E-03	0.85
17063807	*TRBV11-2*	0.30	9.4E-04	0.14	9.3E-02	0.21	0.06	4.5E-04	0.74
17094496	*MIR1299*	0.17	7.7E-02	0.22	2.1E-02	0.20	0.07	3.8E-03	0.91
16842217	*SNORD3D*	0.13	7.0E-02	0.25	2.9E-03	0.18	0.05	8.3E-04	0.74
16818580	*RNA5SP415*	0.16	6.8E-02	0.20	5.1E-02	0.18	0.07	7.9E-03	0.91
16846016	*LRRC37A2*	0.09	3.5E-01	0.24	6.4E-03	0.17	0.06	7.6E-03	0.91
16902656	*RNA5SP103*	0.22	3.0E-03	0.06	5.2E-01	0.16	0.06	5.5E-03	0.91
16975969	*RNU6-310P*	0.20	1.6E-03	0.08	2.5E-01	0.15	0.05	1.6E-03	0.78
16845957	*LRRC37A4P*	0.21	9.3E-03	0.11	1.1E-01	0.15	0.05	3.8E-03	0.91
16788576	*SNORD113-3*	0.14	1.1E-02	0.15	7.2E-03	0.14	0.04	2.0E-04	0.74
16840210	*SLC52A1*	0.19	3.4E-05	0.08	8.6E-02	0.14	0.03	1.7E-05	0.34
16774442	*SMIM2-AS1*	0.14	1.6E-03	0.14	4.8E-03	0.14	0.03	1.9E-05	0.34
16985844	*PMCHL2*	0.11	2.9E-02	0.17	1.4E-02	0.14	0.04	1.3E-03	0.74
16805216	*ASB9P1*	-0.15	1.6E-02	-0.13	3.6E-02	-0.14	0.04	1.4E-03	0.74
17078751	*REXO1L1P*	-0.14	2.3E-03	-0.12	1.0E-01	-0.14	0.04	5.2E-04	0.74
16728518	*FAM86C1*	-0.21	6.6E-03	-0.08	3.1E-01	-0.15	0.06	7.8E-03	0.91
17061129	*RASA4*	-0.20	6.3E-03	-0.09	2.7E-01	-0.16	0.06	5.0E-03	0.91
16836858	*RNU6-446P*	-0.26	3.2E-03	-0.06	5.7E-01	-0.18	0.07	7.1E-03	0.91
16733848	*RP11-326C3.7*	-0.07	4.4E-01	-0.33	9.6E-04	-0.20	0.07	3.8E-03	0.91

Non-typically developing (Non-TD), fold change (FC), standard error (SE);

Several of the 66 nominally differentially-expressed genes between Non-TD and TD cord blood samples, including *DHCR24*, *GNAO1*, *MIR4269*, and *TYMS*, have been previously associated

with genetic variation or gene expression in ASD, although the overlap was not statistically significant (q > 0.05, Fig. S5, Table S2). *MIR4269* was upregulated in Non-TD cord blood ($\log_2$(fold change) = 0.12, p = 7.3E-3), and genetic deficiencies in *MIR4269* have been previously associated with a reduced risk for ASD [40]. When comparing Non-TD differentially-expressed genes in this study to previous expression studies of ASD, *TYMS* and *DHCR24* were downregulated in Non-TD cord blood and also in ASD blood and brain, respectively (*TYMS*: $\log_2$(fold change) = -0.13, p = 7.6E-4; *DHCR24*: $\log_2$(fold change) = -0.11, p = 2.4E-3)[8, 41]. Overlapping differentially expressed genes in both ASD and Non-TD comparisons with TD likely reflect genes that function in general neurodevelopment.

Because genes recently evolved in primates have been hypothesized to play a role in human neurodevelopment, differentially-expressed genes in Non-TD cord blood were assessed for enrichment in recently evolved genes by vertebrate lineages, ranging from tetrapods to homininae using overrepresentation enrichment analysis [45]. Non-TD-associated genes were significantly enriched for genes recently evolved in mammalia, theria, eutheria, boreoeutheria, euarchontoglires, primates, catarrhini, and hominidae, with the greatest enrichment in primate-specific genes (fold enrichment = 7.5, q = 2.1E-5, Fig. 3c, Table S7). Of genes that evolved in primates, *SLC52A1*, *SPANXN5*, and *TRIM49B* were upregulated in Non-TD cord blood, while *FAM86C1*, *RASA4*, *RASA4B*, and *TRIM73* were downregulated (Fig. 3d). In contrast, ASD differentially-expressed genes were not significantly enriched in recently evolved genes from any of the vertebrate lineages (q > 0.05).

After GSEA with all probes ranked by fold change in Non-TD compared to TD subjects, we observed significant enrichment for upregulation of sensory perception gene sets (q < 0.05, Fig. 4a, Table S8). Taste receptor activity (NES = 2.30, q < 1E-4), metabotropic glutamate receptors (NES = 2.21, q = 4.9E-3), and olfactory receptor activity (NES = 1.96, q = 0.018) gene sets were

all upregulated in cord blood from Non-TD subjects. Additionally, gene sets that interact with the compounds quinine (NES = 2.30, q = 2.0E-3) and citric acid (NES = 2.17, q = 2.5E-3) were significantly upregulated, while those interacting with indomethacin (NES = -2.02, q = 0.037) and H2-receptor antagonists (NES = -2.03, q = 0.047) were downregulated. Taste receptor genes included in these gene sets and the top Non-TD-upregulated gene, *TAS2R46*, are located at the 12p13.2 locus, which was also enriched (NES = 2.11, q = 8.3E-3). GSEA was also used to compare Non-TD fold change rankings to genes with cell-type specific expression [38]. Genes upregulated in memory B cells were upregulated in cord blood from Non-TD subjects (NES = 1.69, q = 0.025), while genes upregulated in NK cells and mast cells were downregulated (q < 0.05, Fig. 4b, Table S8). This suggests a shift in the proportions of these cell types in cord blood of Non-TD subjects.

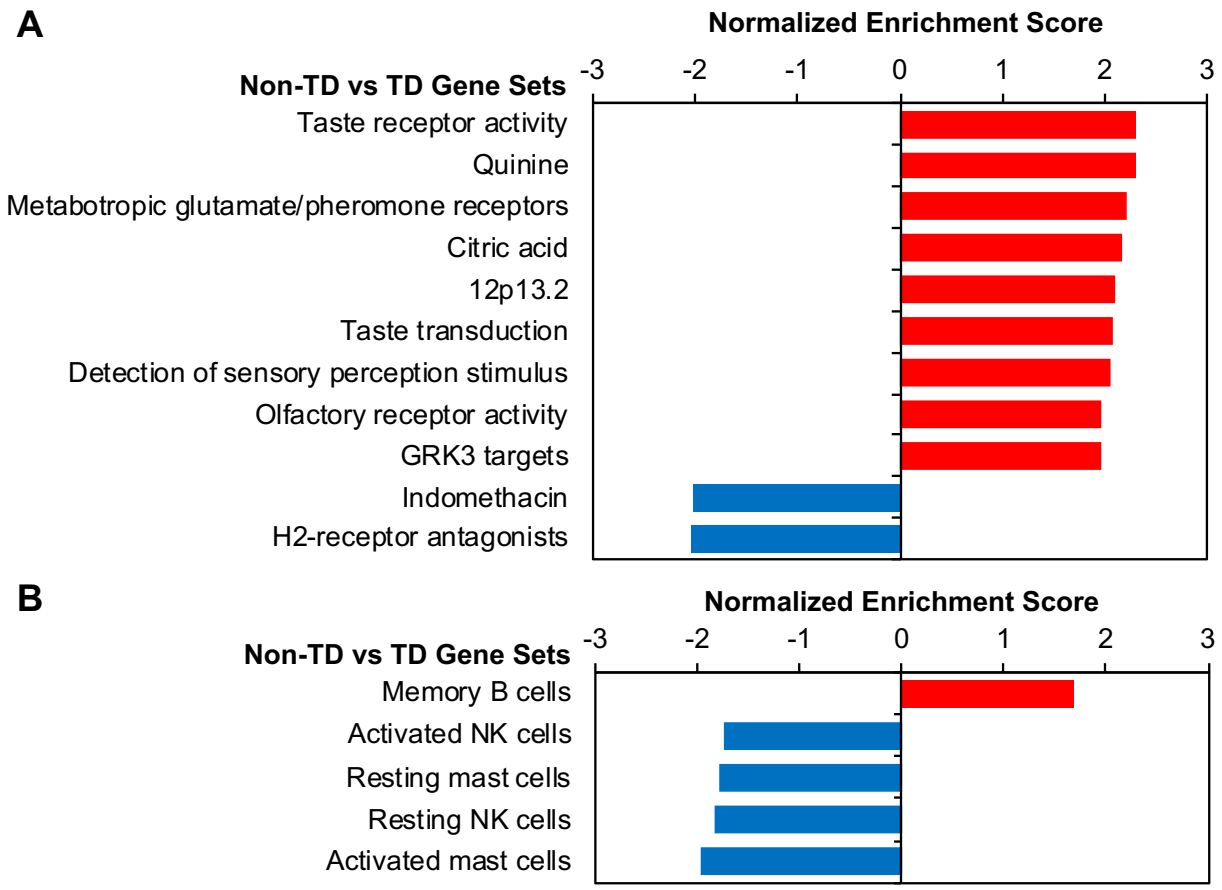

Figure 4. Sensory perception and innate immune cell gene sets are dysregulated in Non-TD subject cord blood. Meta-analysis $\log_2$(fold change) for Non-TD versus TD gene expression was used to rank probes for GSEA with WebGestalt. GSEA assesses whether genes in biologically-defined sets occur toward the top or bottom of a ranked list more than expected by chance. Normalized enrichment score was plotted for significantly enriched gene sets (FDR $q < 0.05$) for (A) default WebGestalt databases, or for (B) blood cell-specific genes.

Comparison of ASD and Non-TD Differentially-Expressed Genes

Eight genes were differentially expressed in both ASD and Non-TD compared to TD subjects, much more than expected by chance (odds ratio = 28.3, p = 1.67E-9, Fig. 5A). Specifically,

IGLV1-40, *LRRC37A4P*, *MIR1299*, *PMCHL2*, and *TRBV11-2* were upregulated, while *RNU4ATAC11P*, *TPTE2P5, and TRIM74* were downregulated in both ASD and Non-TD subjects (Fig. 5B). Further, *LRRC37AP*, *MIR1299*, *PMCHL2*, and *TRBV11-2* were among the top upregulated genes in Non-TD subjects (Fig. 3B). These findings suggest that even though some ASD-associated transcriptional alterations in cord blood are also present in Non-TD subjects, the majority of changes are specific to ASD subjects.

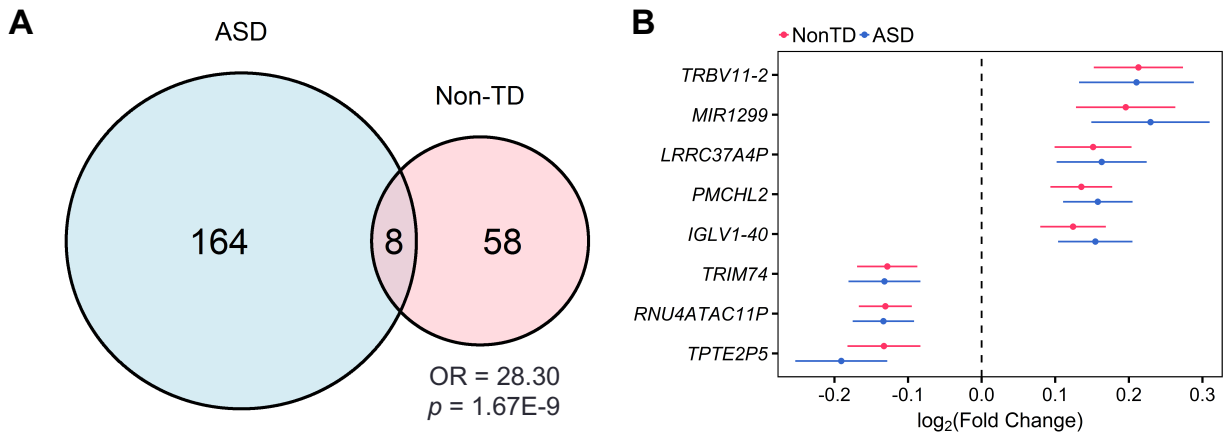

Figure 5. A subset of ASD-associated differentially-expressed genes are also differentially-expressed in Non-TD subjects. (A) Overlap of ASD- and Non-TD-associated differentially-expressed genes from meta-analysis by direction. Significance was assessed by Fisher's Exact Test. (B) Meta-analysis $\log_2$(fold change) in gene expression for ASD- and Non-TD-associated differentially-expressed genes sorted by average $\log_2$(fold change).

Discussion

Perinatal transcriptional alterations in ASD

In this study, we identified gene expression differences in cord blood between high-risk children who went on to develop ASD, were Non-TD, or were TD at 36 months. The results are based on meta-analysis across two high-risk pregnancy cohorts of similar design. Several of these

differentially-expressed genes have been previously associated with neurodevelopmental disorders such as ASD, via genetic variants or previous post-diagnosis expression studies, although overlap with prior findings was not statistically significant. Included in the top upregulated genes in ASD was *TUBB2A*, a component of microtubules. *TUBB2A* is expressed in fetal brain and is mutated in complex cortical dysplasia, a neurodevelopmental disorder that involves seizures [47]. *SLC7A3* and *VSIG4* were also upregulated in ASD and have been previously associated as putative ASD risk genes [39]. Rare missense variants have been found in ASD subjects at *SLC7A3*, which is a cationic amino acid transporter selectively expressed in brain [48]. Nonsense and splice site variants have been found in ASD subjects at *VSIG4*, a complement C3 receptor that functions in pathogen recognition and inhibition of helper T-cell activation [49-51]. Overlap of both genetic and transcriptional association with ASD suggests that dysregulation in these genes may contribute to ASD risk.

Environmental factors are also thought to contribute to ASD risk, especially during the perinatal period, a sensitive window for neurodevelopment [52]. Genes differentially expressed in cord blood from ASD subjects were significantly enriched for functions in xenobiotic metabolism and response to toxic substances. Further, few genes with prior genetic associations to ASD showed differential cord blood expression, which helps elucidate the importance of environmental factors in ASD. Notably, *CYP1A1* was downregulated in ASD cord blood and has been previously found to be transcriptionally regulated in blood by toxicants associated with neurodevelopment, including polychlorinated biphenyls [53-55]. *GCH1*--which is the rate-limiting enzyme in the synthesis of tetrahydrobiopterin, a precursor to folate, dopamine, and serotonin [56]--was also downregulated in cord blood from ASD subjects. *GCH1* expression increases in response to valproic acid, an anti-epileptic drug associated with behavioral deficits in mice and increased risk of autism in humans after *in utero* exposure [57, 58]. Interestingly, *GCH1* is genetically associated with ASD subphenotypes, is downregulated in peripheral blood from ASD

children, and its product tetrahydrobiopterin is decreased in cerebrospinal fluid from ASD subjects [17, 59, 60].

Epigenetic modifications, such as those to DNA and histone protein subunits, are affected by both genetic and environmental factors and are thought to play a role in mediating ASD risk [61, 62]. Immune dysregulation has also been found in children with ASD, and immune cells rely on epigenetic regulation for lineage commitment and cell activation in response to infection [63, 64]. A cluster of histone genes at 6p22.2 was enriched for upregulated genes in ASD cord blood. One of the genes in this cluster was *HIST1H1E*, which encodes H1.4. C-terminal domain truncating mutations in *HIST1H1E* have been associated with autism and Rahman syndrome, a rare disorder that includes intellectual disability and somatic overgrowth [65, 66]. Genes associated with the autoimmune disease systemic lupus erythematosus (SLE) were also upregulated, including histone-encoding, complement pathway, and antigen presentation genes. Epigenetic dysregulation is a feature of SLE, including global DNA hypomethylation, histone H3 and H4 hypoacetylation, and H3K9 hypomethylation in CD4+ T cells [67-69]. Notably, maternal SLE increases risk for ASD in offspring, suggesting an association between these two disorders [70]. Together, this implicates both epigenetic and immune mechanisms in ASD pathobiology.

Perinatal transcriptional alterations in Non-TD

To assess the specificity of cord blood gene expression changes in ASD compared to other neurodevelopmental disorders, this analysis examined transcriptional differences in Non-TD compared to TD subjects across the two studies. The top upregulated gene in Non-TD cord blood was *TAS2R46*, encoding a member of the taste 2 receptor (TAS2R) family, while the top upregulated gene set was taste receptor activity. Upregulated genes in this gene set were primarily other TAS2Rs located at the 12p13.2 locus. TAS2Rs are G-protein coupled receptors (GPCRs) expressed in taste receptor cells and are associated with the perception of bitter taste

[71]. Interestingly, individuals with attention-deficit/hyperactivity disorder and epilepsy have previously been found to perceive bitter tastes more intensely than healthy controls [72, 73]. TAS2Rs are also expressed in blood leukocytes, including NK cells, B cells and T cells, where they may function in chemosensation of foodborne flavor compounds and response to food uptake [74]. Further, TAS2Rs are upregulated in leukocytes from asthma patients and levels of lipopolysaccharide-induced pro-inflammatory cytokines are decreased by TAS2R agonists [75]. Taste receptor upregulation may reflect altered chemosensation in the immune and nervous systems in Non-TD subjects.

Differentially-expressed genes in cord blood from Non-TD subjects were also enriched for genes recently evolved in mammals, and especially for primate-specific genes. These young genes originated at a similar evolutionary time that the neocortex expanded and have biased expression for the fetal neocortex in humans [45]. *SLC52A1* and *RASA4* were among the fetal-biased primate-specific genes that were differentially expressed in Non-TD subjects. *SLC52A1*, which was among the top upregulated genes in Non-TD cord blood, is a transporter involved in the placental uptake of riboflavin, a vitamin important for oxidation-reduction reactions [76]. *RASA4* is a GTPase-activating protein in the Ras signaling pathway that functions in the activation of T cells, mast cells, and macrophages and was also one of the top downregulated genes in cord blood from Non-TD subjects [77-79]. Differential expression of fetal-biased primate-specific genes in Non-TD cord blood may reflect genetic changes that alter expression in multiple tissues.

Children with Non-TD were also observed to have downregulation of *GNAO1*, encoding a G-protein alpha subunit important for neurodevelopment and synaptic signaling. Mutations in *GNAO1* are associated with epileptic encephalopathy, involuntary movements, and intellectual disability [80, 81]. Additionally, missense mutations and downregulation of *GNAO1* in

lymphocytes occur in individuals with schizophrenia [82, 83]. In individuals with ASD, *GNAO1* is upregulated in post-mortem cortex [6]. In murine astrocytes, *Gnao1* and other G-protein effectors are downregulated and intracellular calcium is increased after activation by inflammatory mediators [84]. Further, *GNAO1* is required in mast cells for toll-like receptor 2-mediated pro-inflammatory cytokine release, suggesting *GNAO1* functions in cell signaling in both the nervous and immune systems [85]. Interestingly, we also observed a downregulation of genes highly expressed in mast cells in cord blood from children with Non-TD. Therefore, further investigation into the Non-TD diagnosis group in high-risk families may continue to reveal possible insights into protective gene expression patterns in children resilient to the ASD diagnosis.

Cord blood as a window into transcriptional alterations specific to ASD

Umbilical cord blood gene expression offers a unique snapshot of molecular differences in perinatal life, a critical window for neurodevelopment [86]. Hematopoietic cells in the blood are actively differentiating and responding to environmental cues, such as pathogens and xenobiotics [87, 88]. Epigenetic marks written during this period, which reflect short-term transcriptional activity, have the potential to have long-term effects on gene regulation and cell function [89, 90]. Signals from the immune system cross the blood-brain barrier during gestation and influence the development of the brain [91]. Toxicant exposure during gestation can also impact brain development [92, 93]. In this study, genes involved in toxic substance response, xenobiotic metabolism, and chromatin regulation were altered in cord blood from subjects diagnosed with ASD at 36 months. Transcriptional differences in cord blood from ASD and Non-TD subjects compared to TD subjects were largely independent, with only 8 genes in common. Enriched gene sets associated with Non-TD expression included sensory perception and primate-specific genes and did not overlap with ASD expression gene sets. Further, genes associated with ASD in previous studies of genetic variation and gene expression had few

overlaps with ASD-associated genes in cord blood [6, 8, 39-41]. Instead, cord blood ASD genes likely represent tissue-specific environmentally-responsive genes that may reflect *in utero* exposures and long-term altered neurodevelopmental outcomes.

Conclusions

In the first study to investigate gene expression in cord blood from high-risk newborns later diagnosed with ASD, we identified nominally statistically significant transcriptional alterations specific to ASD, which were enriched for toxic substance response and epigenetic regulation functions. Differentially-expressed genes in ASD had few overlaps with those identified in cord blood from newborns with other non-typical neurodevelopmental outcomes in this high-risk population. Instead, Non-TD-associated genes were enriched for sensory perception functions and primate-specific genes. A strength of this high-risk prospective study design was the observation of gene expression at birth, prior to the onset of symptoms, diagnosis, and treatment. Perinatal life is a critical period for neurodevelopment, where environmental stressors could have long-term impact. Additionally, ASD-associated differential expression was meta-analyzed across two independent studies with harmonized methods and compared with expression changes in other subjects with non-typical neurodevelopment. Finally, cord blood is an accessible tissue that reflects the perinatal environment, and ASD-associated gene expression changes in cord blood may have potential as a predictive biomarker. The limitations of this study include the transcript coverage and measurement precision of the expression microarray platform. Transcripts not covered by a probe on the array were not analyzed in this study, and targeted quantitative analysis in the general population would be needed to validate specific transcriptional changes as ASD risk biomarkers. Additionally, genetic, epigenetic, and environmental variation is known to impact both gene expression and ASD risk, but this was not investigated in this study. Future studies that integrate cord blood gene expression with genetic,

epigenetic, and environmental factors will be important to improve understanding of ASD etiology.

List of abbreviations

Autism Diagnostic Interview-Revised (ADI-R), Autism Diagnostic Observation Schedule (ADOS), autism spectrum disorder (ASD), Early Autism Risk Longitudinal Investigation (EARLI), false discovery rate (FDR), gene set enrichment analysis (GSEA), G-protein coupled receptors (GPCRs), histone deacetylase (HDAC), Markers of Autism Risk in Babies - Learning Early Signs (MARBLES), Mullen Scales of Early Learning (MSEL), natural killer cells (NK cells), non-typically developing (Non-TD), normalized enrichment score (NES), polycomb-repressive complex 2 (PRC2), robust multi-chip average (RMA), Simons Foundation Autism Research Initiative (SFARI), surrogate variable analysis (SVA), systemic lupus erythematosus (SLE), taste 2 receptor (TAS2R), typically developing (TD), University of California (UC),

Declarations

Ethics approval and consent to participate

The UC Davis Institutional Review Board and the State of California Committee for the Protection of Human Subjects approved this study and the MARBLES Study protocols. Human Subjects Institutional Review Boards at each of the four sites in the EARLI Study approved this study and the EARLI Study protocols. Neither data nor specimens were collected until written informed consent was obtained from the parents.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by NIH grants: P01ES011269, R24ES028533, R01ES028089, R01ES020392, R01ES025574, R01ES025531, R01ES017646, U54HD079125, and K12HD051958; EPA STAR grant RD-83329201; and the MIND Institute.

Authors' contributions

CEM and BYP were the lead authors and contributed substantially to data analysis, data interpretation, and drafting the manuscript. KMB contributed critical advice on data analysis methods and data interpretation. JIF and CL contributed to data analysis and interpretation. LAC, CJN, HEV, SO, and IH contributed to study design, as well as subject acquisition, diagnosis, and characterization. JML, RJS, and MDF conceived of the study and contributed substantially to data interpretation and manuscript revision. All authors read and approved the final manuscript.

Acknowledgements

Not applicable.

References

1. Constantino JN, Todorov A, Hilton C, Law P, Zhang Y, Molloy E, et al. Autism recurrence in half siblings: strong support for genetic mechanisms of transmission in ASD. Mol Psychiatry. 2012;18(2):137-8.
2. Sandin S, Lichtenstein P, Kuja-Halkola R, Larsson H, Hultman CM, Reichenberg A. The familial risk of autism. JAMA. 2014;311(17):1770-7.

3. Gaugler T, Klei L, Sanders SJ, Bodea CA, Goldberg AP, Lee AB, et al. Most genetic risk for autism resides with common variation. Nat Genet. 2014;46(8):881-5.
4. Weiner DJ, Wigdor EM, Ripke S, Walters RK, Kosmicki JA, Grove J, et al. Polygenic transmission disequilibrium confirms that common and rare variation act additively to create risk for autism spectrum disorders. Nat Genet. 2017;49(7):978-85.
5. Voineagu I, Wang X, Johnston P, Lowe JK, Tian Y, Horvath S, et al. Transcriptomic analysis of autistic brain reveals convergent molecular pathology. Nature. 2011;474(7351):380-4.
6. Gupta S, Ellis SE, Ashar FN, Moes A, Bader JS, Zhan J, et al. Transcriptome analysis reveals dysregulation of innate immune response genes and neuronal activity-dependent genes in autism. Nat Commun. 2014;5:5748.
7. Ansel A, Rosenzweig JP, Zisman PD, Melamed M, Gesundheit B. Variation in Gene Expression in Autism Spectrum Disorders: An Extensive Review of Transcriptomic Studies. Front Neurosci. 2016;10:601.
8. Tylee DS, Hess JL, Quinn TP, Barve R, Huang H, Zhang-James Y, et al. Blood transcriptomic comparison of individuals with and without autism spectrum disorder: A combined-samples mega-analysis. Am J Med Genet B Neuropsychiatr Genet. 2017;174(3):181-201.
9. Geschwind DH. Autism: many genes, common pathways? Cell. 2008;135(3):391-5.
10. De Rubeis S, He X, Goldberg AP, Poultney CS, Samocha K, Cicek AE, et al. Synaptic, transcriptional and chromatin genes disrupted in autism. Nature. 2014;515(7526):209-15.
11. Oblak AL, Rosene DL, Kemper TL, Bauman ML, Blatt GJ. Altered posterior cingulate cortical cyctoarchitecture, but normal density of neurons and interneurons in the posterior cingulate cortex and fusiform gyrus in autism. Autism Res. 2011;4(3):200-11.
12. Courchesne E, Pierce K. Brain overgrowth in autism during a critical time in development: implications for frontal pyramidal neuron and interneuron development and connectivity. Int J Dev Neurosci. 2005;23(2-3):153-70.

13. Bakulski KM, Feinberg JI, Andrews SV, Yang J, Brown S, S LM, et al. DNA methylation of cord blood cell types: Applications for mixed cell birth studies. Epigenetics. 2016;11(5):354-62.
14. Enstrom AM, Lit L, Onore CE, Gregg JP, Hansen RL, Pessah IN, et al. Altered gene expression and function of peripheral blood natural killer cells in children with autism. Brain Behav Immun. 2009;23(1):124-33.
15. Stamova B, Green PG, Tian Y, Hertz-Picciotto I, Pessah IN, Hansen R, et al. Correlations between gene expression and mercury levels in blood of boys with and without autism. Neurotox Res. 2011;19(1):31-48.
16. Tian Y, Green PG, Stamova B, Hertz-Picciotto I, Pessah IN, Hansen R, et al. Correlations of gene expression with blood lead levels in children with autism compared to typically developing controls. Neurotox Res. 2011;19(1):1-13.
17. Glatt SJ, Tsuang MT, Winn M, Chandler SD, Collins M, Lopez L, et al. Blood-based gene expression signatures of infants and toddlers with autism. J Am Acad Child Adolesc Psychiatry. 2012;51(9):934-44.e2.
18. Gregg JP, Lit L, Baron CA, Hertz-Picciotto I, Walker W, Davis RA, et al. Gene expression changes in children with autism. Genomics. 2008;91(1):22-9.
19. Kong SW, Collins CD, Shimizu-Motohashi Y, Holm IA, Campbell MG, Lee IH, et al. Characteristics and predictive value of blood transcriptome signature in males with autism spectrum disorders. PLoS One. 2012;7(12):e49475.
20. Kong SW, Shimizu-Motohashi Y, Campbell MG, Lee IH, Collins CD, Brewster SJ, et al. Peripheral blood gene expression signature differentiates children with autism from unaffected siblings. Neurogenetics. 2013;14(2):143-52.
21. Hertz-Picciotto I, Schmidt RJ, Walker CK, Bennett DH, Oliver M, Shedd-Wise KM, et al. A Prospective Study of Environmental Exposures and Early Biomarkers in Autism Spectrum

Disorder: Design, Protocols, and Preliminary Data from the MARBLES Study. Environ Health Perspect. 2018;126(11):117004.

22. Newschaffer CJ, Croen LA, Fallin MD, Hertz-Picciotto I, Nguyen DV, Lee NL, et al. Infant siblings and the investigation of autism risk factors. J Neurodev Disord. 2012;4(1):7-1955-4-7.
23. Ozonoff S, Young GS, Carter A, Messinger D, Yirmiya N, Zwaigenbaum L, et al. Recurrence risk for autism spectrum disorders: a Baby Siblings Research Consortium study. Pediatrics. 2011;128(3):e488-95.
24. Lord C, Risi S, Lambrecht L, Cook EH, Jr., Leventhal BL, DiLavore PC, et al. The autism diagnostic observation schedule-generic: a standard measure of social and communication deficits associated with the spectrum of autism. J Autism Dev Disord. 2000;30(3):205-23.
25. Lord C, Rutter M, DiLavore PC, Risi S. The Autism Diagnostic Observation Schedule (ADOS). Los Angeles: Western Psychological Services; 2000.
26. Lord C, Rutter M, Le Couteur A. Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. J Autism Dev Disord. 1994;24(5):659-85.
27. Mullen EM. Scales of Early Learning. Circle Pines, MN: American Guidance Services Inc; 1995.
28. Ozonoff S, Young GS, Belding A, Hill M, Hill A, Hutman T, et al. The broader autism phenotype in infancy: when does it emerge? J Am Acad Child Adolesc Psychiatry. 2014;53(4):398-407 e2.
29. Chawarska K, Shic F, Macari S, Campbell DJ, Brian J, Landa R, et al. 18-month predictors of later outcomes in younger siblings of children with autism spectrum disorder: a baby siblings research consortium study. J Am Acad Child Adolesc Psychiatry. 2014;53(12):1317-27 e1.

30. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics. 2003;4(2):249-64.
31. Kauffmann A, Gentleman R, Huber W. arrayQualityMetrics--a bioconductor package for quality assessment of microarray data. Bioinformatics. 2009;25(3):415-6.
32. Carvalho BS, Irizarry RA. A framework for oligonucleotide microarray preprocessing. Bioinformatics. 2010;26(19):2363-7.
33. Carvalho B. pd.hugene.2.0.st: Platform Design Info for Affymetrix HuGene-2_0-st. R package. 2015;version 3.14.1:https://bioconductor.org/packages/release/data/annotation/html/pd.hugene.2.0.st.html .
34. Leek JT, Storey JD. Capturing heterogeneity in gene expression studies by surrogate variable analysis. PLoS Genet. 2007;3(9):1724-35.
35. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7):e47.
36. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010;26(17):2190-1.
37. Shen L, Sinai M. GeneOverlap: Test and visualize gene overlaps. R package. 2013; version 1.16.0:http://shenlab-sinai.github.io/shenlab-sinai/.
38. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles. Nat Methods. 2015;12(5):453-7.
39. Abrahams BS, Arking DE, Campbell DB, Mefford HC, Morrow EM, Weiss LA, et al. SFARI Gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (ASDs). Mol Autism. 2013;4(1):36.

40. Autism Spectrum Disorders Working Group of The Psychiatric Genomics Consortium. Meta-analysis of GWAS of over 16,000 individuals with autism spectrum disorder highlights a novel locus at 10q24.32 and a significant overlap with schizophrenia. Mol Autism. 2017;8:21.

41. Parikshak NN, Swarup V, Belgard TG, Irimia M, Ramaswami G, Gandal MJ, et al. Genome-wide changes in lncRNA, splicing, and regional gene expression patterns in autism. Nature. 2016;540(7633):423-7.

42. Tylee DS, Espinoza AJ, Hess JL, Tahir MA, McCoy SY, Rim JK, et al. RNA sequencing of transformed lymphoblastoid cells from siblings discordant for autism spectrum disorders reveals transcriptomic and functional alterations: Evidence for sex-specific effects. Autism Res. 2017;10(3):439-55.

43. Durinck S, Spellman PT, Birney E, Huber W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nat Protoc. 2009;4(8):1184-91.

44. Wang J, Duncan D, Shi Z, Zhang B. WEB-based GEne SeT AnaLysis Toolkit (WebGestalt): update 2013. Nucleic Acids Res. 2013;41(Web Server issue):W77-83.

45. Zhang YE, Landback P, Vibranovski MD, Long M. Accelerated recruitment of new brain development genes into the human genome. PLoS Biol. 2011;9(10):e1001179.

46. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005;102(43):15545-50.

47. Cushion TD, Paciorkowski AR, Pilz DT, Mullins JG, Seltzer LE, Marion RW, et al. De novo mutations in the beta-tubulin gene TUBB2A cause simplified gyral patterning and infantile-onset epilepsy. Am J Hum Genet. 2014;94(4):634-41.

48. Nava C, Rupp J, Boissel JP, Mignot C, Rastetter A, Amiet C, et al. Hypomorphic variants of cationic amino acid transporter 3 in males with autism spectrum disorders. Amino Acids. 2015;47(12):2647-58.

49. Lim ET, Raychaudhuri S, Sanders SJ, Stevens C, Sabo A, MacArthur DG, et al. Rare complete knockouts in humans: population distribution and significant role in autism spectrum disorders. Neuron. 2013;77(2):235-42.

50. van Lookeren Campagne M, Verschoor A. Pathogen clearance and immune adherence "revisited": Immuno-regulatory roles for CRIg. Semin Immunol. 2018;37:4-11.

51. Jung K, Seo SK, Choi I. Endogenous VSIG4 negatively regulates the helper T cell-mediated antibody response. Immunol Lett. 2015;165(2):78-83.

52. Kalkbrenner AE, Daniels JL, Chen JC, Poole C, Emch M, Morrissey J. Perinatal exposure to hazardous air pollutants and autism spectrum disorders at age 8. Epidemiology. 2010;21(5):631-41.

53. Leijs MM, Esser A, Amann PM, Schettgen T, Heise R, Fietkau K, et al. Expression of CYP1A1, CYP1B1 and IL-1beta in PBMCs and skin samples of PCB exposed individuals. Sci Total Environ. 2018;642:1429-38.

54. Vorrink SU, Hudachek DR, Domann FE. Epigenetic determinants of CYP1A1 induction by the aryl hydrocarbon receptor agonist 3,3',4,4',5-pentachlorobiphenyl (PCB 126). Int J Mol Sci. 2014;15(8):13916-31.

55. Park HY, Hertz-Picciotto I, Sovcikova E, Kocan A, Drobna B, Trnovec T. Neurodevelopmental toxicity of prenatal polychlorinated biphenyls (PCBs) by chemical structure and activity: a birth cohort study. Environ Health. 2010;9:51.

56. Thony B, Auerbach G, Blau N. Tetrahydrobiopterin biosynthesis, regeneration and functions. Biochem J. 2000;347 Pt 1:1-16.

57. Colleoni S, Galli C, Gaspar JA, Meganathan K, Jagtap S, Hescheler J, et al. A comparative transcriptomic study on the effects of valproic acid on two different hESCs lines in a neural teratogenicity test system. Toxicol Lett. 2014;231(1):38-44.
58. Roullet FI, Lai JK, Foster JA. In utero exposure to valproic acid and autism--a current review of clinical and animal studies. Neurotoxicol Teratol. 2013;36:47-56.
59. Hu VW, Addington A, Hyman A. Novel autism subtype-dependent genetic variants are revealed by quantitative trait and subphenotype association analyses of published GWAS data. PLoS One. 2011;6(4):e19067.
60. Tani Y, Fernell E, Watanabe Y, Kanai T, Langstrom B. Decrease in 6R-5,6,7,8-tetrahydrobiopterin content in cerebrospinal fluid of autistic patients. Neurosci Lett. 1994;181(1-2):169-72.
61. LaSalle JM. Epigenomic strategies at the interface of genetic and environmental risk factors for autism. J Hum Genet. 2013;58(7):396-401.
62. Tordjman S, Somogyi E, Coulon N, Kermarrec S, Cohen D, Bronsard G, et al. Gene x Environment interactions in autism spectrum disorders: role of epigenetic mechanisms. Front Psychiatry. 2014;5:53.
63. Zhao M, Wang Z, Yung S, Lu Q. Epigenetic dynamics in immunity and autoimmunity. Int J Biochem Cell Biol. 2015;67:65-74.
64. Ashwood P, Krakowiak P, Hertz-Picciotto I, Hansen R, Pessah I, Van de Water J. Elevated plasma cytokines in autism spectrum disorders provide evidence of immune dysfunction and are associated with impaired behavioral outcome. Brain Behav Immun. 2011;25(1):40-5.
65. Duffney LJ, Valdez P, Tremblay MW, Cao X, Montgomery S, McConkie-Rosell A, et al. Epigenetics and autism spectrum disorder: A report of an autism case with mutation in H1 linker histone HIST1H1E and literature review. Am J Med Genet B Neuropsychiatr Genet. 2018;177(4):426-33.

66. Tatton-Brown K, Loveday C, Yost S, Clarke M, Ramsay E, Zachariou A, et al. Mutations in Epigenetic Regulation Genes Are a Major Cause of Overgrowth with Intellectual Disability. Am J Hum Genet. 2017;100(5):725-36.

67. Long H, Yin H, Wang L, Gershwin ME, Lu Q. The critical role of epigenetics in systemic lupus erythematosus and autoimmunity. J Autoimmun. 2016;74:118-38.

68. Coit P, Jeffries M, Altorok N, Dozmorov MG, Koelsch KA, Wren JD, et al. Genome-wide DNA methylation study suggests epigenetic accessibility and transcriptional poising of interferon-regulated genes in naive CD4+ T cells from lupus patients. J Autoimmun. 2013;43:78-84.

69. Hu N, Qiu X, Luo Y, Yuan J, Li Y, Lei W, et al. Abnormal histone modification patterns in lupus CD4+ T cells. J Rheumatol. 2008;35(5):804-10.

70. Vinet E, Pineau CA, Clarke AE, Scott S, Fombonne E, Joseph L, et al. Increased Risk of Autism Spectrum Disorders in Children Born to Women With Systemic Lupus Erythematosus: Results From a Large Population-Based Cohort. Arthritis Rheumatol. 2015;67(12):3201-8.

71. Conte C, Ebeling M, Marcuz A, Nef P, Andres-Barquin PJ. Identification and characterization of human taste receptor genes belonging to the TAS2R family. Cytogenet Genome Res. 2002;98(1):45-53.

72. Weiland R, Macht M, Ellgring H, Gross-Lesch S, Lesch KP, Pauli P. Olfactory and gustatory sensitivity in adults with attention-deficit/hyperactivity disorder. Atten Defic Hyperact Disord. 2011;3(1):53-60.

73. Campanella G, Filla A, De Michele G. Smell and taste acuity in epileptic syndromes. Eur Neurol. 1978;17(3):136-41.

74. Malki A, Fiedler J, Fricke K, Ballweg I, Pfaffl MW, Krautwurst D. Class I odorant receptors, TAS1R and TAS2R taste receptors, are markers for subpopulations of circulating leukocytes. J Leukoc Biol. 2015;97(3):533-45.

75. Orsmark-Pietras C, James A, Konradsen JR, Nordlund B, Soderhall C, Pulkkinen V, et al. Transcriptome analysis reveals upregulation of bitter taste receptors in severe asthmatics. Eur Respir J. 2013;42(1):65-78.

76. Yonezawa A, Inui K. Novel riboflavin transporter family RFVT/SLC52: identification, nomenclature, functional characterization and genetic diseases of RFVT/SLC52. Mol Aspects Med. 2013;34(2-3):693-701.

77. Strazza M, Azoulay-Alfaguter I, Dun B, Baquero-Buitrago J, Mor A. CD28 inhibits T cell adhesion by recruiting CAPRI to the plasma membrane. J Immunol. 2015;194(6):2871-7.

78. Nakamura R, Furuno T, Nakanishi M. The plasma membrane shuttling of CAPRI is related to regulation of mast cell activation. Biochem Biophys Res Commun. 2006;347(1):363-8.

79. Zhang J, Guo J, Dzhagalov I, He YW. An essential function for the calcium-promoted Ras inactivator in Fcgamma receptor-mediated phagocytosis. Nat Immunol. 2005;6(9):911-9.

80. Nakamura K, Kodera H, Akita T, Shiina M, Kato M, Hoshino H, et al. De Novo mutations in GNAO1, encoding a Galphao subunit of heterotrimeric G proteins, cause epileptic encephalopathy. Am J Hum Genet. 2013;93(3):496-505.

81. Saitsu H, Fukai R, Ben-Zeev B, Sakai Y, Mimaki M, Okamoto N, et al. Phenotypic spectrum of GNAO1 variants: epileptic encephalopathy to involuntary movements with severe developmental delay. Eur J Hum Genet. 2016;24(1):129-34.

82. Vawter MP, Ferran E, Galke B, Cooper K, Bunney WE, Byerley W. Microarray screening of lymphocyte gene expression differences in a multiplex schizophrenia pedigree. Schizophr Res. 2004;67(1):41-52.

83. Tani M, Mui K, Minami Y, Kiriike N. Association of a GTP-binding protein Go alpha subunit mutation with schizophrenia. Mol Psychiatry. 2001;6(4):359.

84. Hamby ME, Coppola G, Ao Y, Geschwind DH, Khakh BS, Sofroniew MV. Inflammatory mediators alter the astrocyte transcriptome and calcium signaling elicited by multiple G-protein-coupled receptors. J Neurosci. 2012;32(42):14489-510.

85. Jin M, Yu B, Zhang W, Zhang W, Xiao Z, Mao Z, et al. Toll-like receptor 2-mediated MAPKs and NF-kappaB activation requires the GNAO1-dependent pathway in human mast cells. Integr Biol (Camb). 2016;8(9):968-75.

86. Richetto J, Riva MA. Prenatal maternal factors in the development of cognitive impairments in the offspring. J Reprod Immunol. 2014;104-105:20-5.

87. Lura MP, Gorlanova O, Muller L, Proietti E, Vienneau D, Reppucci D, et al. Response of cord blood cells to environmental, hereditary and perinatal factors: A prospective birth cohort study. PLoS One. 2018;13(7):e0200236.

88. Cardenas A, Koestler DC, Houseman EA, Jackson BP, Kile ML, Karagas MR, et al. Differential DNA methylation in umbilical cord blood of infants exposed to mercury and arsenic in utero. Epigenetics. 2015;10(6):508-15.

89. Vineis P, Chatziioannou A, Cunliffe VT, Flanagan JM, Hanson M, Kirsch-Volders M, et al. Epigenetic memory in response to environmental stressors. Faseb j. 2017;31(6):2241-51.

90. Richmond RC, Simpkin AJ, Woodward G, Gaunt TR, Lyttleton O, McArdle WL, et al. Prenatal exposure to maternal smoking and offspring DNA methylation across the lifecourse: findings from the Avon Longitudinal Study of Parents and Children (ALSPAC). Hum Mol Genet. 2015;24(8):2201-17.

91. Bilbo SD, Schwarz JM. The immune system and developmental programming of brain and behavior. Front Neuroendocrinol. 2012;33(3):267-86.

92. Lin CC, Chien CJ, Tsai MS, Hsieh CJ, Hsieh WS, Chen PC. Prenatal phenolic compounds exposure and neurobehavioral development at 2 and 7years of age. Sci Total Environ. 2017;605-606:801-10.

93. Perera F, Phillips DH, Wang Y, Roen E, Herbstman J, Rauh V, et al. Prenatal exposure to polycyclic aromatic hydrocarbons/aromatics, BDNF and child development. Environ Res. 2015;142:602-8.