


RESEARCH

Open Access



Cord blood DNA methylome in newborns later diagnosed with autism spectrum disorder reflects early dysregulation of neurodevelopmental and X-linked genes

Charles E. Mordaunt¹, Julia M. Jianu¹, Benjamin I. Laufer¹, Yihui Zhu¹, Hyeyeon Hwang¹, Keith W. Dunaway¹, Kelly M. Bakulski², Jason I. Feinberg³, Heather E. Volk³, Kristen Lyall⁴, Lisa A. Croen⁵, Craig J. Newschaffer⁶, Sally Ozonoff⁷, Irva Hertz-Picciotto⁸, M. Daniele Fallin³, Rebecca J. Schmidt⁸ and Janine M. LaSalle^{1*} 

Abstract

Background: Autism spectrum disorder (ASD) is a neurodevelopmental disorder with complex heritability and higher prevalence in males. The neonatal epigenome has the potential to reflect past interactions between genetic and environmental factors during early development and influence future health outcomes.

Methods: We performed whole-genome bisulfite sequencing of 152 umbilical cord blood samples from the MARBLES and EARLI high-familial risk prospective cohorts to identify an epigenomic signature of ASD at birth. Samples were split into discovery and replication sets and stratified by sex, and their DNA methylation profiles were tested for differentially methylated regions (DMRs) between ASD and typically developing control cord blood samples. DMRs were mapped to genes and assessed for enrichment in gene function, tissue expression, chromosome location, and overlap with prior ASD studies. DMR coordinates were tested for enrichment in chromatin states and transcription factor binding motifs. Results were compared between discovery and replication sets and between males and females.

Results: We identified DMRs stratified by sex that discriminated ASD from control cord blood samples in discovery and replication sets. At a region level, 7 DMRs in males and 31 DMRs in females replicated across two independent groups of subjects, while 537 DMR genes in males and 1762 DMR genes in females replicated by gene association. These DMR genes were significantly enriched for brain and embryonic expression, X chromosome location, and identification in prior epigenetic studies of ASD in post-mortem brain. In males and females, autosomal ASD DMRs were significantly enriched for promoter and bivalent chromatin states across most cell types, while sex differences were observed for X-linked ASD DMRs. Lastly, these DMRs identified in cord blood were significantly enriched for binding sites of methyl-sensitive transcription factors relevant to fetal brain development.

(Continued on next page)

* Correspondence: jmlasalle@ucdavis.edu

¹Department of Medical Microbiology and Immunology, Genome Center, and MIND Institute, University of California, Davis, CA, USA

Full list of author information is available at the end of the article



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

(Continued from previous page)

Conclusions: At birth, prior to the diagnosis of ASD, a distinct DNA methylation signature was detected in cord blood over regulatory regions and genes relevant to early fetal neurodevelopment. Differential cord methylation in ASD supports the developmental and sex-biased etiology of ASD and provides novel insights for early diagnosis and therapy.

Keywords: Autism spectrum disorder, Neurodevelopment, Umbilical cord blood, Prospective study, Epigenome-wide association study, Epigenetics, Epigenome, DNA methylation, Whole-genome bisulfite sequencing, X chromosome

Background

Autism spectrum disorder (ASD) is a heterogeneous neurodevelopmental disorder that affects 1 in 59 children in the USA [1]. ASD presents as persistent difficulties in social communication and interaction, restricted and repetitive behaviors and interests, as well as sensory sensitivities. Communication deficits can include delayed and monotonous speech, echolalia, poor verbal comprehension, difficulty understanding body language cues, and making eye contact, while behavioral deficits can include stereotyped movements, insistence on routine, fixated interests, and altered sensitivity to sensory input. ASD is currently diagnosed in childhood by one of several standardized scales that include interviews, behavioral observation, and clinical assessment, such as the gold standard Autism Diagnostic Observation Schedule (ADOS) [2, 3]. Clinical management of ASD symptoms consists of both behavioral interventions and pharmacological treatments. Intensive behavioral interventions have been associated with better outcomes in children with ASD, especially for those diagnosed within the first 2–3 years of life as they are the most amenable to behavioral intervention [4].

Risk for ASD is thought to originate from a combination of multiple genetic and environmental factors, as well as gene-environment interactions [5]. Population-based studies have estimated the fraction of variation in ASD risk explained by additive genetic variation to be between 51 and 81%, although this does not account for gene-environment interactions [6]. The genetic architecture of ASD includes both rare variants with strong effects and multiple common variants with weak individual effects [7]. Genes associated with ASD in genetic studies are enriched for pathways affecting neuronal homeostasis and embryonic development and include synaptic neurotransmitter receptors, cytoskeletal proteins, protein degradation factors, and chromatin regulators. However, in keeping with the complexity of ASD, no single genetic variant has been found that accounts for more than 1% of disease liability. Environmental factors are also known to contribute to ASD risk, especially during the prenatal period [8, 9]. During gestation, when rapid cellular proliferation and differentiation is occurring, environmental stressors can have long-lasting effects on behavior [10]. Maternal factors shown to modify ASD risk include environmental toxicants, nutritional factors, fever,

pre-eclampsia, gestational diabetes, and medications [11]. ASD etiology is thought to begin early in development through interactions between genetic and environmental factors that alter neurodevelopmental trajectories [12, 13].

ASD has consistently been found to be more frequent in males compared to females, at about a 3 to 1 ratio [14]. In autistic individuals without intellectual disability, the ratio is increased to 11 to 1 [15]. The sex ratio has been explained in two ways, which are not mutually exclusive: (1) differential ascertainment, where females tend to show more social motivation and ability to hide their social difficulties; and (2) a female protective effect, where females diagnosed with ASD have a larger burden of genetic mutations. In other words, there may be a higher threshold of genetic burden in females required to meet clinical diagnostic criteria. The recurrence rate for both males and females in families containing a female proband is higher than those with only male probands, suggesting a higher genetic load [16, 17]. Females diagnosed with ASD may also have more severe symptoms and comorbidities than their male counterparts [18]. A female protective effect in ASD etiology may be explained by the presence of two copies of chromosome X, which has a disproportionate number of genes involved in neurodevelopment [15]. Through the process of X chromosome inactivation, females are mosaics for mutations in X-linked genes, resulting in a diluted effect of harmful X-linked mutations. Interestingly, some genes on the X chromosome without Y-linked homologs escape X chromosome inactivation, resulting in higher expression in females than in males, and potentially impacting sex-specific susceptibility to ASD [19].

Epigenetic mechanisms have been proposed to account for the sex bias, gene-environment interactions, and developmental origins of ASD etiology [20]. Epigenetic modifications, including DNA methylation, histone post-translational modifications (PTMs), non-coding RNA, and chromatin architecture, are influenced by both genetic and environmental factors and are established dynamically during development to reinforce cell lineage and function [21]. DNA methylation, which primarily occurs as the addition of a methyl group to the 5th carbon of the cytosine in a CpG dinucleotide, is the best understood epigenetic

modification and the most tractable for large human studies. Epigenome-wide association studies (EWAS) have identified locus-specific differential methylation and epigenetic signatures associated with ASD. EWAS in post-mortem brain have identified differential methylation of genes involved in synaptic transmission and the immune response, with a particular enrichment for open chromatin regions and genes important for microglia [22–25]. Similarities in brain epigenetic dysregulation have been found between individuals with idiopathic and syndromic ASDs, including Dup15q and Rett syndrome [24, 26, 27]. Tissues easily accessible in humans, such as placenta, paternal sperm, buccal, and blood, have been used to identify differential methylation in ASD by EWAS [28–32], but few individual loci have been replicated. Previous EWAS have been hampered by study design limitations, including case-control cohorts that may be confounded by reverse causation and microarray-based platforms that cover less than 3% of the CpG sites in the genome [33]. Notably, a previous study from our group applied whole-genome bisulfite sequencing (WGBS) to prospectively collected placenta samples and identified significant differential methylation associated with ASD at *CYP2E1* and *IRS2* [30].

Here, we obtained umbilical cord blood samples from ASD and typically developing (TD) subjects from two high-familial risk prospective cohorts (i.e., cohorts following younger siblings of a child already diagnosed with ASD through that subsequent child's early development) in order to use an accessible tissue at birth before ASD symptoms developed. We used WGBS to assess levels of DNA methylation across more than 20 million CpG sites and sex-stratified the analysis to reveal epigenetic differences specific to males or females. Sex-specific differentially methylated regions (DMRs) distinguishing ASD from TD newborns were analyzed with a systems biology-based approach to identify impacted genes, pathways, transcription factors, and chromatin states. Unlike most previous ASD EWAS, this approach examines methylation before the onset of symptoms, correlates newborn methylation with quantitative clinical measurements at 36 months, covers the majority of CpG sites in the genome, and emphasizes functional genomic regions. As a result of using this approach, we identified sex-specific epigenomic signatures of ASD in umbilical cord blood that replicated in an independent group of subjects.

Methods

Sample population and biosample collection

Markers of Autism Risk in Babies - Learning Early Signs (MARBLES)

The MARBLES study recruited mothers of children receiving services through the California Department of Developmental Services for their diagnosis of ASD;

mothers must have been either planning a pregnancy or already pregnant with another child. Study inclusion criteria were as follows: the mother or father has at least one biological child with ASD; the mother is at least 18 years old; the mother is pregnant; the mother speaks, reads, and understands English at a sufficient level to complete the protocol, the younger sibling will be taught to speak English; and the mother lives within 2.5 h of the Davis/Sacramento region at the time of enrollment. As previously described in depth [34], demographic, lifestyle, environmental, diet, and medical information were collected prospectively through telephone-assisted interviews and mailed questionnaires during pregnancy and the postnatal period. Mothers were provided with umbilical cord blood collection kits prior to delivery. Arrangements were made with obstetricians/midwives and birth hospital labor/delivery staff to ensure proper sample collection and temporary storage. As described below, infants received standardized neurodevelopmental assessments between 6 months and 3 years of age.

Early Autism Risk Longitudinal Investigation (EARLI)

The EARLI study recruited and followed pregnant mothers who had an older child diagnosed with ASD from pregnancy through the first 3 years of life and has been described in detail previously [35]. EARLI families were recruited at four EARLI network sites (Drexel/Children's Hospital of Philadelphia, Johns Hopkins/Kennedy Krieger Institute, Kaiser Permanente Northern California, and University of California, Davis) in three US regions (Southeast Pennsylvania, Northeast Maryland, and Northern California). In addition to having a biological child with ASD confirmed by EARLI study clinicians, inclusion criteria also consisted of being able to communicate in English or Spanish; being 18 years or older; living within 2 h of a study site; and being less than 29 weeks pregnant. EARLI research staff made arrangements with obstetricians/midwives and birth hospital labor/delivery staff to ensure proper umbilical cord blood sample collection and temporary storage. The development of children born into the cohort was closely followed using standardized neurodevelopmental assessments through 3 years of age.

Diagnostic classification

In both the MARBLES and EARLI studies, child development was assessed by trained, reliable examiners. Diagnostic assessments at 36 months included the gold standard ADOS [36], the Autism Diagnostic Interview-Revised [37], conducted with parents, and the Mullen Scales of Early Learning (MSEL) [38], a test of cognitive, language, and motor development. Based on a previously published algorithm that uses ADOS and MSEL scores [39, 40], children included in the study were classified into one of three exclusive outcome groups: ASD, TD,

or non-typically developing; however, the analyses for this paper were limited to samples from children with ASD and TD outcomes. Children with ASD outcomes had scores over the ADOS cutoff and met DSM-5 criteria for ASD. Children with TD outcomes had all MSEL scores within 2 standard deviations (SDs) of the average and no more than one MSEL subscale 1.5 SDs below the normative mean and scores on the ADOS at least three points below the ASD cutoff.

Demographic characteristics

In both studies, demographic information was collected prospectively throughout gestation and the postnatal period with in-person and telephone-assisted interviews and mailed questionnaires. Demographic characteristics were tested for differences relative to diagnostic outcome across all subjects using Fisher's exact test for categorical variables and one-way ANOVA for continuous variables. *P* values were adjusted for the total number of demographic variables using the false discovery rate (FDR) method [41].

WGBS sample processing

DNA was extracted from whole umbilical cord blood with the Gentra Puregene Blood kit (Qiagen, Hilden, Germany) and quantified with the Qubit dsDNA HS Assay kit (Thermo Fisher Scientific, Waltham, MA, USA). DNA was bisulfite converted with the EZ DNA Methylation Lightning kit (Zymo, Irvine, CA, USA). Sodium bisulfite treatment converts all unmethylated cytosine to uracil residues, which are detected as thymine after library preparation [42]. WGBS libraries were prepared from 100 ng of bisulfite-converted DNA using the TruSeq DNA Methylation kit (Illumina, San Diego, CA, USA) with indexed PCR primers and a 14-cycle PCR program. Libraries for the discovery sample set were sequenced at 2 per lane with 150 base pair paired-end reads and spiked-in PhiX DNA on the HiSeq X (Illumina, San Diego, CA, USA) by Novogene (Sacramento, CA, USA). Samples included in the discovery set were obtained from 74 males (TD *n* = 39, ASD *n* = 35) and 32 females (TD *n* = 17, ASD *n* = 15) in the MARBLES and EARLI studies (MARBLES *n* = 42, EARLI *n* = 64). Libraries for the replication sample set were sequenced over two lanes indexed at 4 samples per lane with 100 base pair single-end reads and spiked-in PhiX DNA on the HiSeq 4000 (Illumina, San Diego, CA, USA) by the Vincent J. Coates Genomics Sequencing Laboratory at University of California, Berkeley. Samples included in the replication set were obtained from 38 males (TD *n* = 17, ASD *n* = 21) and 8 females (TD *n* = 3, ASD *n* = 5) in the MARBLES study.

WGBS read alignment and quality control

Sequencing reads were preprocessed, aligned to the human genome, and converted to CpG methylation count matrices with CpG_Me (v1.0, [43–45]) using the appropriate

pipeline for single- or paired-end reads. After Illumina quality filtering of raw fastq files, reads were trimmed to remove adapters and both 5' and 3' methylation bias. Trimmed reads were screened for contaminating genomes, aligned to the hg38 human reference genome assembly, and filtered for PCR duplicates. Counts of CpG methylation at all covered sites were extracted to generate Bismark cytosine methylation reports. Comprehensive quality control reports were examined for each sample, and libraries with incomplete bisulfite conversion were excluded, as measured by CHH methylation greater than 2%. Sex was confirmed for each sample based on coverage of the X and Y chromosomes. The CpG_Me workflow incorporates the Trim Galore! (v0.4.5, RRID:SCR_011847, [46]), Bismark (v0.19.1, RRID:SCR_005604, [44]), Bowtie 2 (v2.3.4.1, RRID:SCR_005476, [47]), FastQ Screen (v0.11.4, RRID:SCR_000141, [48]), SAMtools (v1.8, RRID:SCR_002105, [49]), and MultiQC (v1.5, RRID:SCR_014982, [50]) packages.

Global DNA methylation analysis and covariate associations

Global CpG methylation for each sample was extracted from Bismark count matrices as the total number of methylated CpG counts divided by the total CpG counts across all chromosomes. Global methylation was compared with behavioral, demographic, and technical variables using linear regression, stratified by sex, and examined both within and pooled across sequencing platforms. Linear models included adjustment for PCR duplicates and also sequencing platform when combined. Continuous variables were converted to SD before linear regression. *P* values were adjusted for the number of variables using the FDR method.

Tiled window methylation and principal component analysis (PCA)

DNA methylation at 10-kb windows tiled across the genome was obtained using a custom perl script (Window_permeth_readcentric.pl [51]). Windows were filtered for those with at least 1 read over at least 10 CpGs in all samples, and then percent methylation for each sample was calculated as the number of methylated CpG counts divided by the total CpG counts within the window. PCA was performed using the prcomp() function in the stats R package with centering to zero and scaling to unit variance, and then visualized using the ggbiplot R package (v0.55, [52]) with the ellipse indicating the 95% confidence limit for each group.

Cell-type deconvolution

Cell-type proportions were deconvoluted from WGBS methylation data based on a umbilical cord blood reference panel, which defined cell-type-specific CpG sites in B cells, CD4+ and CD8+ T cells, granulocytes, monocytes,

natural killer cells, and nucleated red blood cells using the Infinium HumanMethylation450 BeadChip array [53]. Percent methylation was extracted from WGBS Bismark methylation count matrices at cell-type-specific loci covered in at least 90% of samples, which included between 76 and 85 CpG sites per cell type. Missing values were imputed with the mean percent methylation for that locus across all samples. Percent methylation data, along with model parameters defined in the reference panel, were input into the `projectCellType()` function in the `minfi` R package (v1.28.4, RRID:SCR_012830, [54]) to estimate cell-type proportions, which were scaled to 100% for each sample. We empirically determined that the use of this `minfi` approach was more accurate at determining cell proportions than the `methylCC` R package (v1.0.0, [55]), which estimates cell composition in a platform-agnostic manner by identifying completely methylated or unmethylated cell-type-specific regions and modeling these as latent states, although there was a high correlation between both methods (Pearson's $r = 0.85$, $p = 1.3E-292$), suggesting these estimates are robust to the specific computational method.

DMR and differentially methylated block (DMB) identification

DMRs were identified between ASD and TD subjects in only males, only females, and all subjects with adjustment for sex. DMBs were identified between ASD and TD subjects within sex only. No other adjustment covariates were included. The same comparisons for DMRs and DMBs were done in the discovery and replication sets. DMRs and DMBs were identified using the `DMRICH` (v1.0, [43, 56]), `dmrseq` (v1.2.3, [57]), and `bsseq` (v1.18.0, RRID:SCR_001072, [58]) R packages. The beta-binomial distribution, variance in percent methylation, and spatial correlations inherent in WGBS data can be appropriately modeled using a generalized least squares regression model with a nested autoregressive correlated error structure as implemented in the `dmrseq` package [57, 59]. In this approach, candidate regions are identified based on consistent differences in mean methylation between groups, and region-level statistics are estimated which account for coverage, mean methylation, and correlation between CpGs. With this approach, the number of statistical tests is not equal to the CpGs examined (20 million), but rather the number of candidate regions (about 20,000). These region statistics are then compared to a permutation-generated pooled null distribution to calculate an empirical p value. Permutations are used to calculate the null distribution of test statistics because it may differ depending on the particular samples or tissues used and cannot be assumed a priori.

In this study, Bismark count matrices were filtered for CpGs covered by at least 1 read in 50% of samples in both groups, but if less than 10 samples were present in

a group, the threshold was increased to 1 read in all samples. DMRs were identified with the `dmrseq()` function from the `dmrseq` package using the default parameters, except the single CpG methylation difference coefficient cutoff was set to 0.05 and the minimum number of CpGs was set to 3. DMRs were called if the permutation-based p value was less than 0.05. FDR-adjusted p values were calculated within each model. DMBs were also identified with the `dmrseq()` function using the default parameters for blocks, except the single CpG methylation difference coefficient cutoff was set to 0.01 and the minimum number of CpGs was set to 3. As described in the `dmrseq` package vignette [57], the default parameters for DMBs differ from those for DMRs in that the minimum width for DMBs is 5 kb and the maximum gap between CpGs in a DMB is also 5 kb. Additionally, the smoothing span window is widened by setting the minimum CpGs in a smoothing window to 500, the width of the window to 50 kb, and the maximum gap between CpGs in the same smoothing cluster to 1 Mb. Background regions were defined as the genomic locations where it is possible to identify a DMR, which were those regions in the coverage-filtered Bismark count matrix with at least 3 CpG sites less than 1 kb apart.

DMR hierarchical clustering and PCA

The ability of DMRs to distinguish between ASD and TD subjects was tested through clustering subjects based on percent methylation in DMRs using hierarchical clustering and PCA. In both approaches, percent methylation in DMRs was extracted from Bismark methylation count matrices as the number of methylated CpG counts divided by the total CpG counts in that DMR. For hierarchical clustering, the mean percent methylation in a DMR across all subjects was subtracted from the percent methylation in a DMR for each subject. Both subjects and DMRs were clustered using the Euclidean distance and Ward's agglomeration method. For PCA, missing values were imputed using the `imputePCA()` function in the `missMDA` R package (v1.16, [60]). PCA was performed as above using the `prcomp()` function in the `stats` R package with centering to zero and scaling to unit variance, and then visualized using the `ggbiplot` R package (v0.55, [52]) with the ellipse indicating the 95% confidence limit for each group.

DMR-covariate associations

Percent methylation in DMRs was extracted from Bismark methylation count matrices as the number of methylated CpG counts divided by the total CpG counts in that DMR. Continuous variables were converted to SD, and methylation at each DMR was compared with behavioral, demographic, and technical variables using linear regression without any adjustments. P values were adjusted for

the number of variables and DMRs within each comparison using the FDR method.

Computational validation of DMRs

DMRs from all comparisons conducted within sex were examined for validation by recalling them using the `BSmooth.tstat()` and `dmrFinder()` functions from the `bsseq` R package (v1.18.0, RRID:SCR_001072, [58]), which uses a different statistical approach. In this method, percent methylation values for individual CpGs are smoothed to incorporate information from neighboring loci. *T*-tests are conducted at each CpG site to compare the two groups, and adjacent CpGs with *p* values less than 0.05 are grouped into DMRs. Default parameters were used for `BSmooth.tstat()`, except the variance was estimated assuming it was the same for both groups. Default parameters were also used for `dmrFinder()`, except the *t*-statistic cutoff was set to the value where $p = 0.05$ in a two-sided *t*-test with $n - 2$ degrees of freedom, according to the `qt()` function. DMRs were further filtered for at least 3 CpGs, average methylation difference greater than 0.05, and inverse density of at most 300. The genomic locations of the DMRs from the two different methods and with the same direction of methylation difference were overlapped. The significance of the overlap between the DMR sets was tested using a permutation-based test implemented with the `regioneR` R package (v1.14.0, [61]). Both sets of DMRs were redefined as the set of background regions containing a DMR for that set. Permutation-based *p* values were calculated by comparing the true overlap to a null distribution of overlaps between 10,000 length-matched region sets randomly sampled from the background. This type of permutation test has the advantage of taking into account the complexity of the genome and not requiring assumptions about an underlying statistical model. Because DMRs only represent a tiny portion of the genome (< 0.2%), very few overlaps are expected by chance.

Technical validation of DMRs with bisulfite pyrosequencing

DNA extraction and sodium bisulfite treatment were performed in the same manner as above for WGBS sample processing. Bisulfite pyrosequencing primers were designed using `PyroMark Assay Design` software (Qiagen, Hilden, Germany) to target CpGs within DMRs from the male replication comparison. Bisulfite-converted DNA from each sample was amplified in triplicate for each primer set using the `PyroMark PCR kit` (Qiagen, Hilden, Germany) according to the manufacturer's instructions, except using 50 ng of bisulfite-converted DNA and 40 PCR cycles. Each PCR product was checked with gel electrophoresis for specific amplification. Pyrosequencing was conducted using `PyroMark Gold Q96 SQA Reagents` (Qiagen, Hilden, Germany) with the `PyroMark Q96 instrument` (Qiagen, Hilden,

Germany) according to the manufacturer's instructions. Mean percent methylation for each region was compared to diagnosis outcome using linear regression and compared to WGBS methylation using Pearson's *r*.

DMR replication by region overlap, differential methylation, and gene overlap

DMRs identified within sex between ASD and TD subjects in the discovery sample set were examined for replication in the independent replication sample set using three different methods: region overlap, differential methylation, and gene overlap. In the region overlap method, DMRs were identified separately for samples in the discovery and replication sets and the genomic locations of DMRs with the same direction of methylation difference were overlapped. Significance of the overlap between the discovery and replication DMRs was tested using a permutation-based test implemented with the `regioneR` R package (v1.14.0, [61]), the same as with the computational validation of the DMRs, except the background regions from the discovery and replication sets were intersected to generate a consensus background.

In the differential methylation replication method, percent methylation at DMRs identified in discovery samples was first extracted from Bismark count matrices for both discovery and replication samples. DMRs not covered in more than 50% of either group in the replication sample set were excluded. Percent methylation at each DMR was compared with diagnosis outcome separately in discovery and replication samples using linear regression without adjustments, and replicated DMRs were called as those with the same direction and an unadjusted *p* value less than 0.05 in both sample sets.

In the gene overlap method, genes were first annotated to DMRs using the same approach employed by the `Genomic Regions Enrichment of Annotations Tool` (GREAT) [62], but adapted for the hg38 assembly. In this strategy, every gene in the NCBI RefSeq database is given a basal regulatory domain that extends 5 kb upstream and 1 kb downstream of the transcription start site (TSS), regardless of other genes. The basal regulatory domain is then extended upstream and downstream of the TSS to the nearest gene's proximal domain or 1 Mb, whichever is closer, to define the broad regulatory domain. Finally, a gene is assigned to a DMR if it overlaps the DMR's broad regulatory domain. All genes annotated to discovery DMRs for that sex were overlapped with all genes annotated to replication DMRs for that sex. The significance of DMR gene overlap between the discovery and replication sets was tested using Fisher's exact test implemented with the `GeneOverlap` R package (v1.18.0, [63]), and compared to genes annotated to both sets of background regions. Replicated DMR genes were defined as those annotated to both discovery and replication

DMRs for that sex. Replicated DMR genes in males were tested for significant overlap with replicated DMR genes in females as above. DMB genes were also examined for replication with the gene overlap method. To test the dependence of DMR gene overlap on the width of the regulatory domain, DMRs were mapped to genes using the same approach, but limiting the regulatory domain to a maximum of 20 kb upstream and downstream of the TSS. Overlap between discovery and replication comparisons was tested for significance as above. These conservatively mapped DMR genes were also tested for functional enrichment and overlap with other ASD studies.

DMR replication through machine learning

Predictive models were performed separately on male and female DMRs, using methylation at discovery ASD DMRs in discovery samples as the training set and methylation at these same regions in replication samples as the testing set. Batch effects in the initial training and testing sets from different sequencing platforms were corrected for using the ComBat() function in the sva R package (v3.34.0, RRID:SCR_012836, [64]). For male samples, the parameters of the function were specified to use a model matrix consisting of the variable of interest (Diagnosis) and the initial training set as the reference batch, which adjusted the mean and variance of the initial testing set to match the initial training set. Batch effects for female samples were corrected similarly as in the male samples, except the model matrix consisted of the variable of interest (Diagnosis) and covariates (CpG methylation percentage, CHG methylation percentage, CHH methylation percentage, trimmed percentage, aligned percentage, and duplicated percentage) to use as adjustment variables. To select the best performing model using the caret R package (v6.0.85, [65]), k -fold cross-validation ($k=20$ for males and $k=5$ for females) with the random forest model was applied on batch-corrected male and female training sets 3 times with different mtry parameters, which specify the number of variables randomly sampled at each tree node split. For both males and females, the model with mtry = 2 was selected as the final model to predict the diagnosis class of the samples in the respective batch-corrected testing sets.

Gene functional enrichment

ASD DMR gene functional enrichment was assessed using the Database for Annotation, Visualization, and Integrated Discovery (DAVID, v6.8, RRID:SCR_001881, [66]) and the accompanying RDAVIDWebService R package (v1.20.0, [67]). For both discovery and replication sets, enrichments of genes annotated to sex-stratified ASD DMRs were compared with genes annotated to background regions, and these were input into DAVID as NCBI Entrez ID numbers. P values were adjusted using the Benjamini

method and significantly enriched terms were called as those with q -values less than 0.05. Enriched terms were defined as replicated if they were significantly enriched in both discovery and replication DMR genes. Functional enrichment of DMB genes was assessed as above. Functional enrichments of all X chromosome genes were also done similarly, with X-linked RefSeq genes compared to all RefSeq genes using their Entrez IDs.

Autism gene enrichment

For both discovery and replication sets, enrichment of sex-stratified ASD DMR genes with those identified in previous genomic, epigenomic, and transcriptional studies of ASD was examined. Gene lists were sourced from the reported hits of each paper [7, 13, 22–31, 68–89]. Reported regions from epigenomic studies were converted to the hg38 assembly using the liftOver() function from the rtracklayer R package (v1.42.2, [90]), and genes were annotated using the same approach used in GREAT [62]. After converting to Entrez IDs, overlap significance was tested with Fisher's exact test using the GeneOverlap R package (v1.18.0, [63]) and compared to genes annotated to background regions. P values were adjusted for the number of gene lists using the FDR method, and significant overlaps were called as those with q -values less than 0.05. Replicated overlaps were identified as the gene lists that significantly overlapped with both discovery and replication DMR genes.

CpG context, ChIP-seq, and ChromHMM chromatin state enrichments

For both discovery and replication sets, sex-stratified ASD DMRs were assessed for region-based enrichment with CpG context, histone PTM ChIP-seq peaks, and ChromHMM-defined chromatin states using the Locus Overlap Analysis (LOLA) R package (v1.12.0, [91]). CpG context maps were obtained from the annotatr R package (v1.8.0, [92]), while histone ChIP-seq peaks and chromatin states were obtained from the Roadmap Epigenomics Project (RRID:SCR_008924, [93]). As recommended in the LOLA documentation [91], enrichments were conducted with DMRs redefined as the set of background regions containing a DMR compared to the set of all background regions. For appropriate visualization in heatmaps, overlaps with less than 5 regions were excluded, and infinite odds ratios and p values were replaced with the maximum value for that DMR set. P values were adjusted with the FDR method, and top enriched regulatory regions were defined as those where the q -value was less than 0.05 and both the odds ratio and $-\log_{10}(q\text{-value})$ were at least the median value for more than half of the 127 cell types examined. Significance of differential enrichment of X chromosome compared to autosome ASD DMRs was assessed by

paired two-sided t -test of odds ratios for each cell type for that regulatory region. P values were adjusted for the number of different regulatory regions with the FDR method, and significant differential enrichment was called if the q value was less than 0.05.

Transcription factor motif enrichment

For both discovery and replication sets, enrichment for known transcription factor binding site motif sequences in sex-stratified ASD DMRs was examined using the Hypergeometric Optimization of Motif EnRichment (HOMER) tool (v4.10, RRID:SCR_010881, [94]). After redefining DMRs as the subset of background regions containing a DMR, DMRs and background regions were input into findMotifsGenome.pl with the default parameters, except the region size was set to given, sequences were normalized for percent CpG content, and the number of randomly sampled background regions was increased to 100,000. Fold enrichment was calculated as the percent of DMR sequences with the motif divided by the percent of background sequences with the motif. P values were adjusted for the number of known motifs tested using the FDR method. Top enriched motifs were defined as those with a q -value less than 0.05 and both fold enrichment and $-\log_{10}(q\text{-value})$ in at least the top quartile for that DMR set. Replicated motifs were those identified as top enriched in both discovery and replication DMR sets for that direction.

Replicated transcription factor motif analysis

Transcription factors that replicated as among the top enriched in sex-stratified ASD DMRs across the discovery and replication sample sets were investigated for motif enrichment in fetal brain chromatin states, expression in fetal brain, and methylation sensitivity. To test for motif enrichment in chromatin states in fetal brain, genomic locations for replicated motif sequences were first determined with scanMotifGenomeWide.pl from HOMER (v4.10, RRID:SCR_010881, [94]). Background regions were defined as the set of regions in the genome where any of the known motifs tested in HOMER were present. Enrichment in male or female fetal brain chromatin states from the Roadmap Epigenomics Project (RRID:SCR_008924, [93]) was tested using the LOLA R package (v1.12.0, [91]), with replicated motif locations redefined as a subset of the background regions containing a motif. P values were adjusted using the FDR method, and top enriched chromatin states were those with a q -value less than 0.05, and with both odds ratio and $-\log_{10}(q\text{-value})$ in at least the top quartile for that tissue type. The top ranked chromatin state for each motif was the top enriched state with the lowest mean rank of odds ratio and p value. RNA-seq expression data of replicated transcription factors in fetal brain were

obtained from the Allen BrainSpan Atlas of the Developing Human Brain (RRID:SCR_008083, [95]). Displayed data were derived from the dorsolateral prefrontal cortex from one male (ID# 12820) and one female (ID# 12834) donor at 13 weeks post conception. RNA-seq expression data of replicated X-linked DMR genes were also derived from this source. Methylation sensitivity data were obtained from MethMotif (v1.3, [96]), and counts for each motif corresponding to unmethylated (less than 10%), partially methylated (10–90%), and methylated (greater than 90%) loci were summed and divided by the total counts for that motif to determine the proportion of binding sites in each methylation category.

Post hoc power analysis

A conservative power analysis was performed in order to estimate the number of samples and detectable methylation differences for future studies, using the ssize.twoSamp() function from the ssize.fdr R package [97]. The FDR was controlled at 0.05, and the 90th percentile of the pooled SD was estimated from methylation in background regions and weighted by group sample size. The proportion of true negatives (π_0) was estimated from p values of background region methylation by diagnosis. The methylation difference for 0.8 power was calculated, as well as the estimated power to detect methylation differences of 5%, 10%, 15%, or 20%.

Results

Study subject characteristics in relation to neurodevelopmental outcome at 36 months

Subjects in this study were from the MARBLES (TD $n = 44$, ASD $n = 44$) and EARLI (TD $n = 32$, ASD $n = 32$) high-familial risk prospective cohorts, and they include those with an ASD outcome and those determined to be TD at 36 months. By definition of the high-familial risk cohort design, all subjects have an older sibling with ASD. Additionally, due to the sex bias in ASD, the majority (74%) of the subjects are males (Additional file 2: Table S1, S2). The ADOS assessment scale is a criteria in the diagnostic algorithm for ASD (see “Methods”), and as such the ASD subjects exhibit significantly higher severity on the ADOS comparison score [3] (ASD mean = 6.6, TD mean = 1.1, $q = 8.0E-62$). The MSEL [98] measures cognitive functioning and was used to exclude subjects with intellectual disability from those classified with TD; thus, ASD subjects also had lower performance on the MSEL early learning composite score (ASD mean = 80, TD mean = 112, $q = 4.2E-21$) when compared to TD subjects. Mothers of ASD subjects differed from mothers of TD subjects on a few characteristics. Specifically, mothers of ASD subjects had altered educational attainment ($p = 0.02$) including increased attainment of at most an associate’s degree (ASD = 24%,

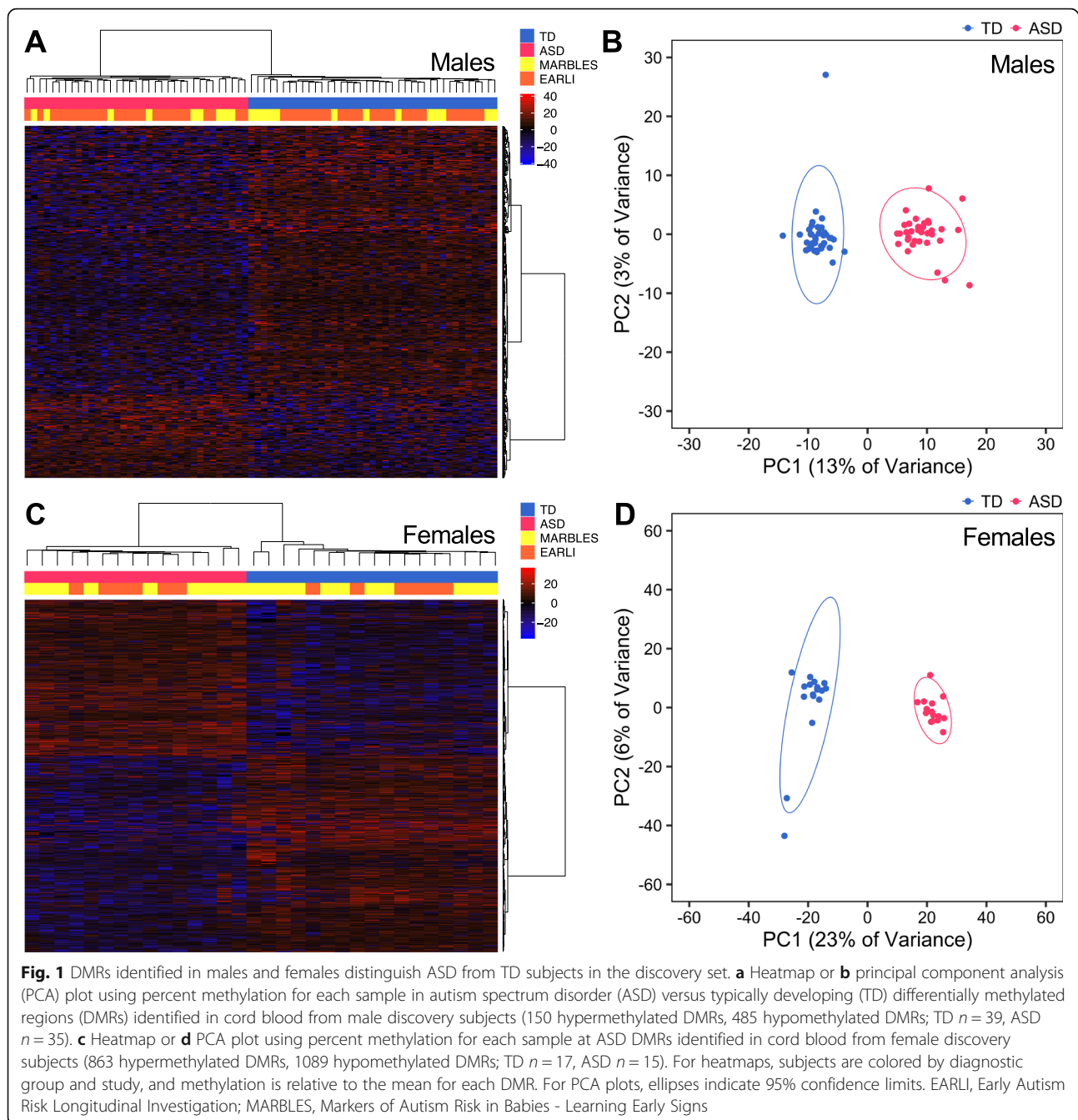
TD = 11%) and decreased attainment of at most a bachelor's degree (ASD = 20%, TD = 43%). They also were less likely to own their home (ASD = 50%, TD = 70%, $p = 0.01$), had a higher pre-pregnancy body mass index (ASD mean = 29, TD mean = 27, $p = 0.04$), and were more likely to smoke during pregnancy (ASD = 11%, TD = 1%, $p = 0.03$).

Lower global DNA methylation in umbilical cord blood from males later diagnosed with ASD corresponds with increased nucleated red blood cells

ASD subjects were found to have lower global CpG methylation than TD subjects (Additional file 2: Table S1). Because this has been found previously in blood from children with ASD [99], we investigated this association in more detail and tested the hypothesis that there could be sex differences in global CpG methylation. Specifically, analyses were performed stratified by sex, both within and pooled across sequencing platforms, and included adjustment for duplicate reads and sequencing platform. Only males with ASD were significantly hypomethylated compared to males with TD (pooled estimate = -0.55% methylation, $q = 0.02$), while females with ASD had similar global methylation as females with TD (pooled estimate = $+0.14\%$ methylation, $q = 0.92$; Additional file 1: Fig. S1A, S2A, Additional file 2: Table S3). Furthermore, scores on the MSEL were positively associated with global methylation only in males, where male subjects with lower scores also had lower global methylation (pooled estimate = $+0.30\%$ methylation per SD, $q = 0.02$; Additional file 1: Fig. S1). Other examined demographic and technical covariates were not associated with global methylation, with the exception of the proportion of nucleated red blood cells (nRBCs) estimated based on percent methylation at cell-type-specific loci [53] (Additional file 1: Fig. S1A, S2D, Additional file 2: Table S3). In both males and females, nRBCs were negatively associated with global methylation (pooled males estimate = -0.71% methylation per SD, $q = 8.6E-18$; pooled females estimate = -0.71% methylation per SD, $q = 7.0E-4$). Similar to the pattern observed with global methylation, nRBCs were increased only in males with ASD compared to males with TD (pooled estimate = $+2.65\%$ nRBCs, $q = 0.046$; Additional file 1: Fig. S2B, Additional file 2: Table S3), and were also negatively associated with MSEL scores only in males (pooled estimate = -1.21% nRBCs per SD, $q = 0.046$; Additional file 1: Fig. S2C). Importantly, when nRBCs were added as an adjustment covariate, global methylation was no longer associated with ASD diagnosis in males (pooled estimate = -0.18% methylation, $p = 0.19$; Additional file 2: Table S3). These findings suggest that an increased proportion of nRBCs in whole cord blood, specifically in males later diagnosed with ASD, can explain their lower observed global methylation levels.

Region-specific differential methylation patterns in umbilical cord blood distinguish males and females later diagnosed with ASD from TD controls

Because DNA methylation, like ASD etiology, is influenced by both genetic and environmental factors during prenatal life, we hypothesized that umbilical cord blood DNA from newborns who later develop ASD would exhibit DMRs compared to those with TD. To test this hypothesis, we analyzed cord blood WGBS data from the MARBLES and EARLI studies for DMRs using an inference and permutation-based statistical approach that is conducive to identifying broad epigenomic signatures of multiple gene regulatory loci (methylation differences of more than 5% over regional clusters of at least 3 CpGs less than 1 kb apart and permutation-based p values less than 0.05) [24]. This approach offers a systems biology perspective at the level of gene pathways and regulatory elements, rather than focusing on a narrow set of high-confidence genetic loci. These data were all generated on the HiSeq X sequencing platform and represent the "discovery set" (males TD $n = 39$, ASD $n = 35$; females TD $n = 17$, ASD $n = 15$). Because of the expected sex differences in DNA methylation patterns due to X chromosome inactivation and the previously observed sex differences in subjects with ASD [15], we performed a sex-stratified analysis to preserve sex-specific differential methylation. In males, 150 hypermethylated (ASD greater than TD) and 485 hypomethylated (ASD less than TD) DMRs associated with ASD were identified, and methylation in these regions clustered subjects distinctly by outcome (Fig. 1a,b, Additional file 2: Table S4). Similarly, ASD DMRs were identified in females, including 863 hypermethylated and 1089 hypomethylated DMRs that could distinguish between subjects with ASD and TD outcome (Fig. 1c,d, Additional file 2: Table S5). When methylation levels within male or female ASD DMRs were compared with subject characteristics, they were specifically associated with autism severity and MSEL cognitive scores but not other demographic and technical variables (Additional file 1: Fig. S3, Additional file 2: Table S4, S5). Notably, the majority of ASD DMRs were not associated with the proportion of nRBCs, which was a driver of global methylation ($p > 0.05$, males: 77% of DMRs not associated, females: 94% of DMRs not associated). Similarly, when nRBC proportion was included as an adjustment covariate to identify DMRs, a significant majority of ASD DMRs were maintained (Additional file 2: Table S6). Likewise, methylation at most DMRs was associated with ASD outcome whether or not the proportions of all cord blood cell types were included as adjustment covariates, suggesting the identified ASD DMRs are robust to variation in cell composition (Additional file 2: Table S7). An alternate approach of combining males and females, with sex included as an adjustment covariate, revealed similar results



but fewer ASD-associated DMRs (Additional file 1: Fig. S4, Additional file 2: Table S8).

To validate our finding that DMRs between ASD and TD subjects are present in cord blood, we replicated the WGBS on an independent group of subjects in the MARBLES study using a different sequencing platform. These subjects were analyzed separately since the sequencing platform showed a larger influence on methylation than MARBLES/EARLI study (Additional file 1: Fig. S5). Data from these subjects were all generated on the HiSeq 4000 sequencing platform and represent the

“replication set” (males TD $n = 17$, ASD $n = 21$; females TD $n = 3$, ASD $n = 5$). Similar to the discovery set, sex-stratified DMRs were identified in the replication set that could specifically cluster ASD versus TD subjects and were more strongly associated with autism severity and cognition than other variables (4650 DMRs in males, 8728 DMRs in females, Additional file 1: Fig. S6, S7, Additional file 2: Table S9, S10). Again, most ASD DMRs in males or females were not associated with the proportion of nRBCs ($p > 0.05$, males: 75% of DMRs not associated, females: 98% of DMRs not associated) and

the inclusion of the estimated nRBC proportion as an adjustment covariate resulted in the significant overlap with a majority of ASD DMRs (Additional file 2: Table S6). DMRs were also identified in the replication set with a combined sex-adjusted approach, which similarly resulted in fewer DMRs than the sex-stratified analysis (Additional file 1: Fig. S8, Additional file 2: Table S11). To computationally validate the detected DMRs, we re-did all comparisons with a different statistical approach, as implemented in the *bsseq* R package [58]. DMRs in all comparisons significantly overlapped between the two methods, suggesting these ASD DMRs identified in cord blood are not an artifact of the computational approach (permutation test, z -score > 35 and $p < 1.0E-4$ for all; Additional file 2: Table S12). Interestingly, when we compared the ability to replicate the WGBS-derived ASD DMRs with Infinium HumanMethylation450 or MethylationEPIC arrays, 82% and 84% of DMRs from the male discovery and replication sets, respectively, did not overlap even one array probe (Additional file 1: Fig. S9). Similarly, 71% and 69% of DMRs identified in females from the discovery and replication sets, respectively, were not covered on either array, reinforcing the utility of the WGBS approach. To validate the detected DMRs with a targeted technique, we assayed 4 DMRs in a subset of samples using bisulfite pyrosequencing and observed a significant correlation with percent methylation data obtained from WGBS at all 4 regions (Pearson's $r > 0.42$ and $q < 0.05$ for all, Additional file 1: Fig. S10, Additional file 2: Table S13). Together, these results demonstrate the discovery and technical replication of novel DMRs in cord blood associated with ASD outcome at 36 months in two high-familial risk cohorts.

ASD DMRs in umbilical cord blood replicate across independent groups of subjects

After confirming that we could identify ASD DMRs in cord blood, we next hypothesized that specific regions are consistently differentially methylated in both discovery and replication sets of newborns later diagnosed with ASD. Testing for the replication of individual DMRs was undertaken with three approaches: region overlap, differential methylation, and gene overlap. When DMRs identified in males from the discovery set were overlapped by genomic location with those found in males from the replication set, 4 hypermethylated DMRs and 3 hypomethylated DMRs were present in both (out of 635 discovery DMRs), which was more than expected by chance (permutation test, hypermethylated z -score = 8.1, $p < 1.0E-4$; hypomethylated z -score = 6.1, $p < 1.0E-4$; Additional file 2: Table S14, S15). In females, 7 hypermethylated and 24 hypomethylated DMRs were identified in both subject sets (out of 1954 discovery DMRs), also a significant overlap (permutation test, hypermethylated z -score = 14.2, $p <$

$1.0E-4$; hypomethylated z -score = 23.7, $p < 1.0E-4$). As a reference, very few overlaps are expected by random chance, because DMRs represent less than 0.2% of the genome. To determine whether ASD DMRs detected in the discovery set showed the same directional differential methylation in the replication set, we compared the percent methylation over each of these regions to ASD outcome in both sample sets. In males, 15 out of 635 DMRs in the discovery set were differentially methylated similarly in the replication set, while 23 out of 1954 discovery DMRs were directionally similar in females from both sample sets ($p < 0.05$; Additional file 2: Table S16). We also compared methylation at diagnosis DMRs from the discovery set to symptom severity within ASD subjects and identified 1 out of 635 DMRs in males and 3 out of 1954 DMRs in females with consistent correlation in both sample sets ($p < 0.05$; Additional file 1: Fig. S11, Additional file 2: Table S17).

As an alternative method for testing replication of ASD DMRs, we used a machine learning approach that included k -fold cross-validation with a random forest model trained and tested separately on males and females (Additional file 1: Fig. S12). In each case, the model was trained using methylation at discovery DMRs for samples in the discovery set and tested on methylation at these same regions for replication samples. The random forest model trained using the male samples in the discovery set correctly classified 11 ASD and 14 TD samples and misclassified 13 samples in the male replication set, labelling 3 TD samples as ASD and 10 ASD samples as TD (Additional file 1: Fig. S13). The model performed moderately with an observed accuracy of 65.8% (25/38). The balanced accuracy, a more reliable metric for imbalanced classes calculated as the average of sensitivity and specificity, was slightly higher at 67.4%. The expected accuracy, or accuracy by random chance, was 48.6%. The kappa statistic, which ranges from -1 to 1 and indicates how a model performed compared to random chance, was 0.334 and in the fair agreement range. The random forest model trained using discovery set female samples correctly classified 4 ASD and 3 TD samples and misclassified 1 ASD sample as TD in the female replication set (Additional file 1: Fig. S14). The model performed well with an observed accuracy of 87.5% (7/8), a balanced accuracy of 90%, an expected accuracy of 50%, and a kappa statistic of 0.750, which indicates substantial agreement. Although the machine learning models for males and females performed moderately and well, respectively, these are only preliminary results to provide evidence for the validity of ASD DMRs. Limitations of the small sample sizes were accommodated through batch effect correction and k -fold cross-validation, but replication with larger sample sets

would be required to obtain more robust and comprehensive models.

Because the functional output of the genome is the transcription of genes, we compared genes annotated to ASD DMRs in each subject set, using broad regulatory domains previously mapped to genes by GREAT (described in “Methods”), which we subsequently refer to as “ASD DMR genes.” For males, a significant enrichment of 537 out of 949 DMR genes in the discovery set were also in the replication set (odds ratio = 5.1, $p = 1.2E-126$; Table 1, Additional file 2: Table S18). Similarly, for females, DMR genes in the discovery set significantly overlapped with those in the replication set, with 1762 out of 2912 present in both (odds ratio = 2.9, $p = 3.6E-153$). Replicated DMR genes in males also significantly overlapped with those in females, with 162 genes replicated in both sexes (odds ratio = 6.2, $p = 4.0E-60$). However, the majority of replicated ASD DMR genes were specific to males or females (375/537 male-specific, 1600/1762 female-specific). DMR genes also replicated in the combined sex-adjusted analysis, but most of the genes identified when stratifying for sex were unique (odds ratio = 5.5, $p = 1.9E-45$; Additional file 1: Fig. S15). Additionally, DMR genes in the discovery set still significantly overlapped with those in the replication set for both males and females, when using a more conservative gene mapping approach that limited DMRs to no more than 20 kb from the TSS (Additional file 2: Table S19). These findings suggest that reproducible ASD DMRs are present in cord blood, and these are not dependent on the particular group of subjects or technical approaches.

Genes in ASD DMBs replicate between independent groups of subjects and are enriched for ASD DMR genes, cadherins, and developmental genes

Because of the differences detected in global CpG methylation, we also tested the hypothesis that large DMBs (defined as regions more than 5 kb in length with greater than 1% methylation difference in ASD versus TD) are present in cord blood DNA from individuals with ASD. DMBs were indeed present in males and

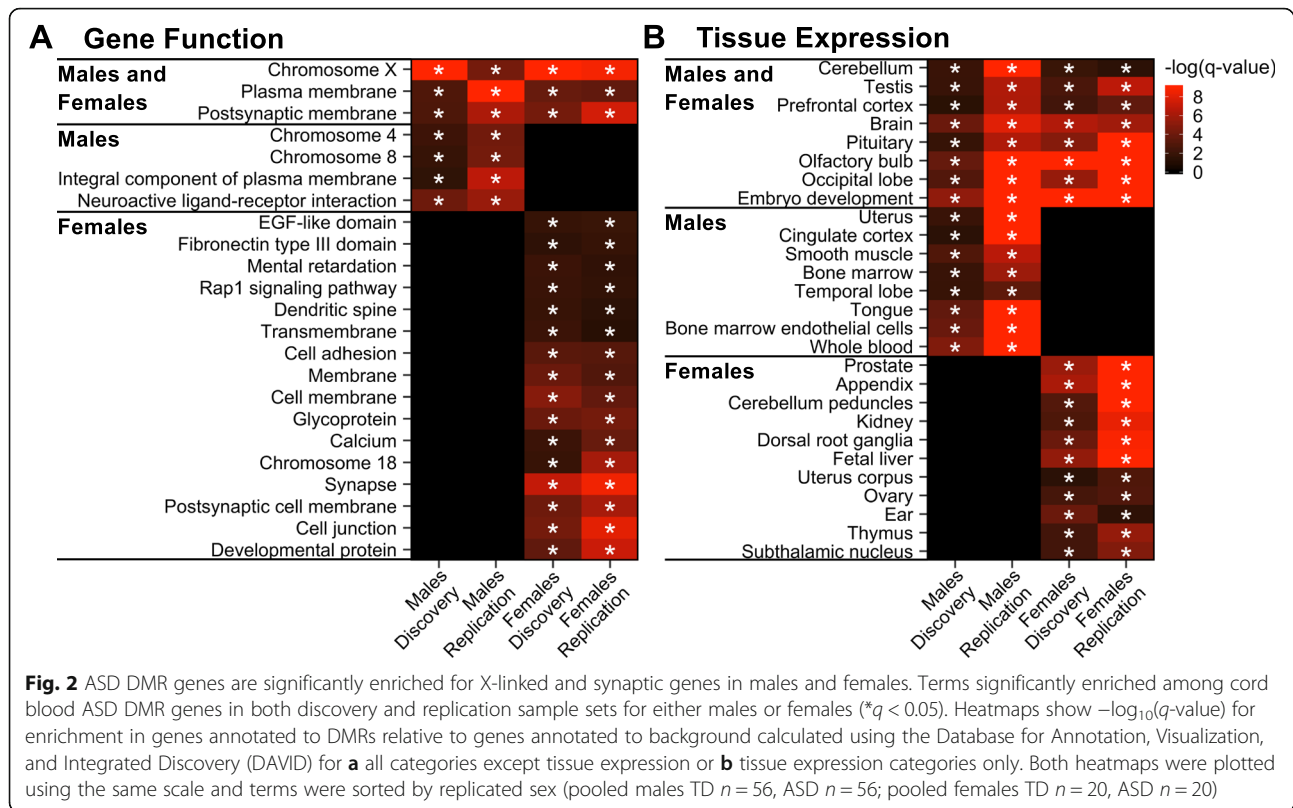
females from both the discovery and replication sets (Additional file 2: Table S20–S23). A comparison of genes annotated to DMBs in the discovery and replication sets revealed a significant overlap of 58 genes in males and 23 genes in females that were present across both sets (male odds ratio = 3.7, $p = 1.7E-13$; female odds ratio = 2.6, $p = 1.7E-4$; Additional file 2: Table S18, S24). A significant enrichment of replicated cord blood ASD DMR genes was observed in these replicated DMB genes, including 20 genes in males and 12 genes in females (male odds ratio = 25.8, $p = 1.3E-19$; female odds ratio = 15.1, $p = 6.1E-9$; Additional file 1: Fig. S16, Additional file 2: Table S18, S24). In males and females from both subject sets, DMB genes were significantly enriched for chromosome 5 location, cell adhesion and calcium signaling functions, and expression during embryonic development ($q < 0.05$; Additional file 1: Fig. S17, Additional file 3: Table S25). Male DMB genes were specifically enriched for expression in brain and bone marrow endothelial cells, while female DMB genes were especially enriched for cadherin genes in both subject sets ($q < 0.05$). These findings suggest that ASD DMBs are present in cord blood, and they impact some of the same genes and pathways as ASD DMRs, with a particular impact on cadherins.

Cord blood ASD DMR genes are enriched for neurodevelopmental genes on the X chromosome that are epigenetically dysregulated in ASD brain

To test the hypothesis that ASD DMR genes identified in cord blood are functionally relevant to ASD etiology, we examined these genes for enrichment in predefined gene sets including pathways, chromosomal location, tissue expression, and previous genome-wide studies of ASD. For both the discovery and replication sets, ASD DMR genes detected in males and females were significantly enriched for X chromosome location and functions localized to the postsynaptic membrane ($q < 0.05$; Fig. 2, Additional file 3: Table S26). ASD DMR genes identified from male and female cord blood were also enriched for expression in embryonic development, multiple brain regions, pituitary, and testes ($q < 0.05$). Genes

Table 1 ASD DMR genes replicate in males and females in independent sample sets

Gene set 1	Genes	Gene set 2	Genes	Overlap	% Overlap	Odds ratio	<i>p</i> value
Discovery	949	Replication	5429	537	56.6	5.13	1.2E–126
Males (<i>n</i> = 74)		Males (<i>n</i> = 38)					
Discovery	2912	Replication	9496	1762	60.5	2.86	3.6E–153
Females (<i>n</i> = 32)		Females (<i>n</i> = 8)					
Replicated	537	Replicated	1762	162	30.2	6.20	4.0E–60
Males (All chrs)		Females (All chrs)					
Replicated	55	Replicated	173	21	38.2	3.48	4.6E–5
Males (chrX)		Females (chrX)					



that replicated in both males and females and are expressed in brain, at the postsynaptic membrane, and during embryonic development include *GABRA2*, *GABRG1*, *GRIA3*, *GRIK2*, *LRRC4C*, *LRRTM1*, *LRRTM4*, *KCNC2*, and *ZC4H2*. In males but not females, cord blood ASD DMR genes were enriched for locations on chromosomes 4 and 8, expression in cingulate cortex, temporal lobe, and blood, and for functions in neuroactive ligand-receptor interactions ($q < 0.05$). In contrast, female DMR genes identified in cord blood were enriched for genes associated with chromosome 18, mental retardation, dendritic spines, and calcium, as well as dorsal root ganglia and subthalamic nucleus expression ($q < 0.05$). Similar enrichment results were also obtained using a more conservative gene mapping approach (Additional file 3: Table S27). ASD DMR genes in cord blood were enriched for X chromosome genes in both males and females and these significantly overlapped, with 21 genes in common (odds ratio = 3.5, $p = 4.6E-5$; Table 1, Fig. 3). However, most replicated genes on the X chromosome were specific for either males or females, with more DMR genes detected in females (34/55 male-specific, 152/173 female-specific). Interestingly, all genes on the X chromosome (not just DMR genes) are enriched for brain and embryonic expression, as well as mental retardation and autism, compared to all genes in the genome ($q < 0.05$; Additional file 1: Fig. S18, Additional file 3: Table S28). Accordingly, many replicated X-linked ASD DMR genes

are expressed in fetal brain (Additional file 1: Fig. S19, S20).

To further examine relevance to ASD etiology, we overlapped cord blood ASD DMR genes with previously reported gene sets from genetic, epigenetic, and transcriptional studies of ASD samples. The majority (77%) of the replicated X chromosome cord blood ASD DMR genes in males or females were identified in previous studies of ASD (Fig. 3). Genome-wide, cord blood ASD DMR genes in males and females from both the discovery and replication sets were significantly enriched for genes identified in previous studies examining DNA methylation (by both array-based and WGBS approaches) in the cingulate cortex, prefrontal cortex, and temporal cortex, and also histone 3 lysine 27 (H3K27) acetylation in prefrontal cortex and cerebellum from subjects with ASD ($q < 0.05$; Fig. 4 shows replicated enrichments, Additional file 1: Fig. S21 shows all comparisons, and Additional file 3: Table S29 shows statistics, genes, and sources). Notably, 16 genes that replicated in both males and females from our WGBS analyses were also identified in at least four epigenetic studies of ASD post-mortem brain, including *CHST15*, *CPXM2*, *FAM49A*, *FAM155B*, *KDR*, *LINC01491*, *LINC01515*, *LOC100506585*, *LOC100996664*, *LOC101928441*, *MIR378C*, *RBFOX1*, *RBMS3-AS1*, *ST6GAL2*, *TGFBR2*, and *TIMP3*. Many of the genes also identified in previous epigenomic studies of ASD are located on the X

Males and Females	<i>AFF2</i> ••	<i>KLHL13</i> ••	<i>LOC101928437</i>	<i>PCDH11X</i> •	<i>TMEM47</i> •	<i>ZIC3</i> ••
	<i>CPXCR1</i> •	<i>LINC00269</i> •	<i>LOC389906</i> •	<i>PPP4R3CP</i>	<i>XACT</i>	
	<i>FAM155B</i> •	<i>LINC00889</i>	<i>MAGEB16</i>	<i>SPANXN1</i> •	<i>XRCC6P5</i> •	
	<i>GRIA3</i> ••	<i>LOC101928201</i> •	<i>MIR3672</i> •	<i>TDGF1P3</i> •	<i>ZC4H2</i>	
Males Only	<i>BRWD3</i> •	<i>FAM133A</i> ••	<i>IGSF1</i> •	<i>MAGEC2</i> •	<i>SPANXN4</i> •	<i>TGIF2LX</i> •
	<i>CAPN6</i> •	<i>FAM47A</i> ••	<i>ITM2A</i> ••	<i>MIR548M</i>	<i>SPIN4</i> •	<i>UBE2NL</i> •
	<i>CD40LG</i> •	<i>FAM47B</i>	<i>LINC01201</i> •	<i>MIR891A</i> •	<i>SRPX</i> •	<i>VCX3A</i>
	<i>CLDN34</i> •	<i>FUNDC1</i> •	<i>LOC100126447</i> •	<i>SH3BGRL</i>	<i>STS</i> ••	<i>ZXDA</i> •
	<i>CXorf51B</i>	<i>GLRA2</i>	<i>LOC101927501</i>	<i>SHROOM2</i> ••	<i>TBX22</i> •	
	<i>EFHC2</i> •	<i>HMG5</i> ••	<i>LOC101928495</i>	<i>SPANXD</i>	<i>TEX13C</i>	
Females Only	<i>ACTRT1</i>	<i>DIAPH2-AS1</i> •	<i>IDS</i> •••	<i>MID1</i> •••	<i>NLGN3</i> ••	<i>RPA4</i> •
	<i>ADGRG4</i> •	<i>DLG3-AS1</i> •	<i>IL1RAPL1</i> •••	<i>MID1IP1-AS1</i> •	<i>NR0B1</i> ••	<i>SASH3</i> •
	<i>AMER1</i> •	<i>DOCK11</i> •	<i>IL1RAPL2</i> •	<i>MID2</i> ••	<i>NUP62CL</i> •	<i>SCML1</i> •
	<i>ANOS1</i> •	<i>DUSP9</i> ••	<i>KIAA1210</i> •	<i>MIR1468</i>	<i>OCRL</i> ••	<i>SCML2</i> •
	<i>APLN</i> •	<i>EDA</i> ••	<i>KLF8</i> •	<i>MIR1587</i> •	<i>OPN1LW</i>	<i>SLC7A3</i> •••
	<i>AR</i> •••	<i>EDA2R</i> •	<i>KLHL4</i> •	<i>MIR2114</i> •	<i>PAGE1</i>	<i>SLITRK2</i> ••
	<i>ARHGAP6</i> ••	<i>EFNB1</i> ••	<i>LAMP2</i> •	<i>MIR223</i>	<i>PCDH19</i> ••	<i>SMS</i> •••
	<i>BCOR</i> •••	<i>ELF4</i> ••	<i>LANCL3</i>	<i>MIR325HG</i>	<i>PCSK1N</i> •••	<i>SNX12</i> •
	<i>BCORL1</i> •	<i>ERAS</i> ••	<i>LINC01186</i> •	<i>MIR3915</i>	<i>PGAM4</i> •	<i>SPANXN3</i> ••
	<i>BRDTP1</i> •	<i>ESX1</i> •	<i>LINC01281</i> •	<i>MIR3937</i> •	<i>PGK1</i> •••	<i>SSX1</i>
	<i>BRS3</i> •	<i>FAM47C</i> •	<i>LOC100421746</i>	<i>MIR424</i> •	<i>PGRMC1</i> ••	<i>STARD8</i> •
	<i>CASK</i> •••	<i>FAM58A</i> •	<i>LOC100506790</i>	<i>MIR4536-2</i>	<i>PHEX</i> •	<i>SUPT20HL2</i>
	<i>CCDC22</i> •	<i>FAM9A</i>	<i>LOC101928335</i> •	<i>MIR4666B</i> •	<i>PHEX-AS1</i> •	<i>TBL1X</i> •••
	<i>CDKL5</i> ••	<i>FGF16</i>	<i>LOC101928402</i> •	<i>MIR513C</i> •	<i>PIR-FIGF</i>	<i>TEX11</i> •
	<i>CFAP47</i>	<i>FIGF</i> •	<i>LOC101928627</i> •	<i>MIR514A3</i>	<i>PJA1</i> ••	<i>UBQLN2</i> ••
	<i>CHMP1B2P</i>	<i>FMR1</i> ••	<i>LOC401585</i> •	<i>MIR548AX</i> •	<i>PLP2</i> •	<i>UPRT</i>
	<i>CITED1</i> •	<i>FRMPD3</i> •	<i>LUZP4</i> •	<i>MIR548I4</i>	<i>PLS3</i> •	<i>USP51</i> •
	<i>CLCN4</i> ••	<i>FRMPD4</i> ••	<i>MAGEA4</i>	<i>MIR6134</i>	<i>POU3F4</i> •	<i>VCX</i> •
	<i>CNGA2</i>	<i>FTHL17</i> •	<i>MAGEA6</i>	<i>MIR676</i> •	<i>PPP1R2P9</i> •	<i>VENTXP1</i>
	<i>CUL4B</i> •••	<i>GLUD2</i> •	<i>MAGEB18</i>	<i>MOSPD1</i>	<i>PRDX4</i> •	<i>VMA21</i>
	<i>CXorf21</i> •	<i>GPR101</i> •	<i>MAGEB2</i> •	<i>MSL3</i>	<i>PRR32</i> •	<i>ZDHHC15</i>
	<i>CXorf51A</i> •	<i>GPR143</i> •	<i>MAGED2</i> ••	<i>MSN</i> ••	<i>PRRG1</i> ••	<i>ZNF645</i>
	<i>DACH2</i>	<i>GPR82</i> •	<i>MAMLD1</i> •	<i>MTM1</i> ••	<i>PTCHD1</i> •••	
	<i>DCAF8L1</i>	<i>HDAC8</i> ••	<i>MAOA</i> ••	<i>MTMR1</i> ••	<i>PUDP</i> •	
	<i>DDX53</i> •	<i>HDX</i> •	<i>MAP7D3</i> •••	<i>MTMR8</i>	<i>RAI2</i> •	
	<i>DIAPH2</i> ••	<i>HTR2C</i>	<i>MECP2</i> •••	<i>MUM1L1</i> ••	<i>RLIM</i> ••	

• Genetic ASD Study • Epigenetic ASD Study • Gene Expression ASD Study

Fig. 3 Replicated sex-independent, male-specific, and female-specific X-linked DMR genes and their overlap with previous ASD studies. Genes on the X chromosome that were annotated to cord blood ASD DMR genes in both discovery and replication sample sets for males and/or females. Dots indicate overlap with at least 1 genetic, epigenetic, or gene expression study of ASD

chromosome and expressed in fetal brain, including *GRIA3*, *AFF2*, and *KLHL13*, which replicated in males and females; *SHROOM2*, *ZXDA*, and *HMG5*, which were male-specific; and *MECP2*, *FMR1*, and *PCSKIN*, which were female-specific (Additional file 1: Fig. S22-S24). Male and female DMR genes also significantly overlapped those identified as differentially methylated in prefrontal cortex from the syndromic ASDs Dup15q and Rett syndromes, and in sperm from fathers of toddlers displaying ASD-like traits in EARLI ($q < 0.05$). Female but not male ASD DMR genes were significantly enriched for dysregulated gene sets from other ASD DNA methylation studies in brain, but also in MARBLES placenta, lymphoblast cell lines, and

newborn blood spots ($q < 0.05$). Additionally, female ASD DMR genes significantly overlapped with known ASD risk genes and those associated with altered histone 3 lysine 4 (H3K4) trimethylation in neurons from ASD subjects ($q < 0.05$). Similar overlap with genes identified in other ASD studies was also found using a more conservative gene mapping approach (Additional file 3: Table S30). Reflecting their specificity, ASD DMR genes did not significantly overlap in both discovery and replication sets with risk genes for Alzheimer's disease or with genes that were differentially methylated in Alzheimer's cortex.

In striking contrast to the strong concordance across epigenetic ASD studies using different marks and

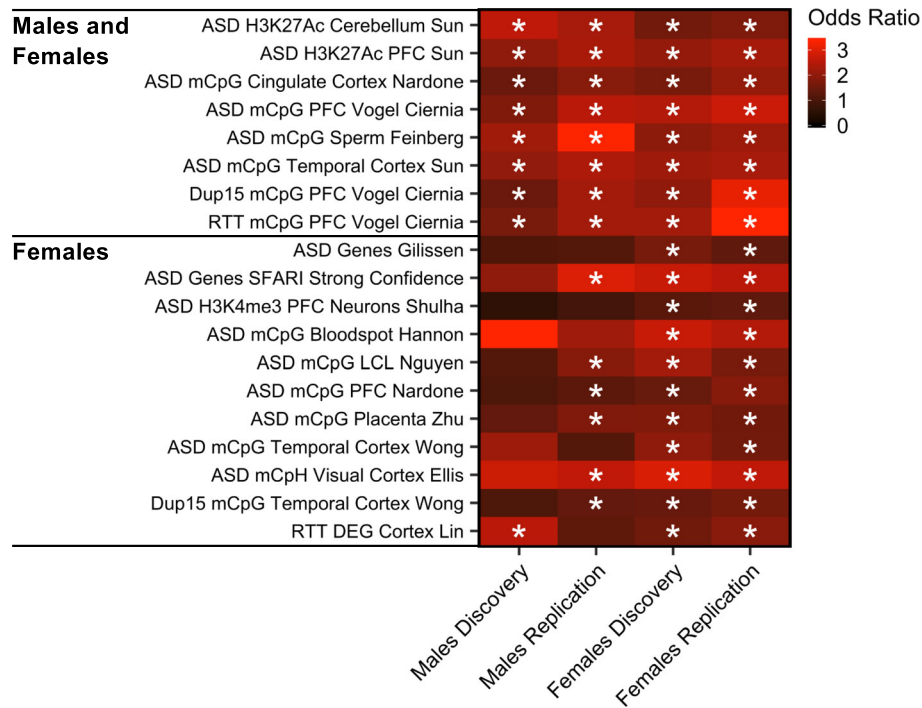


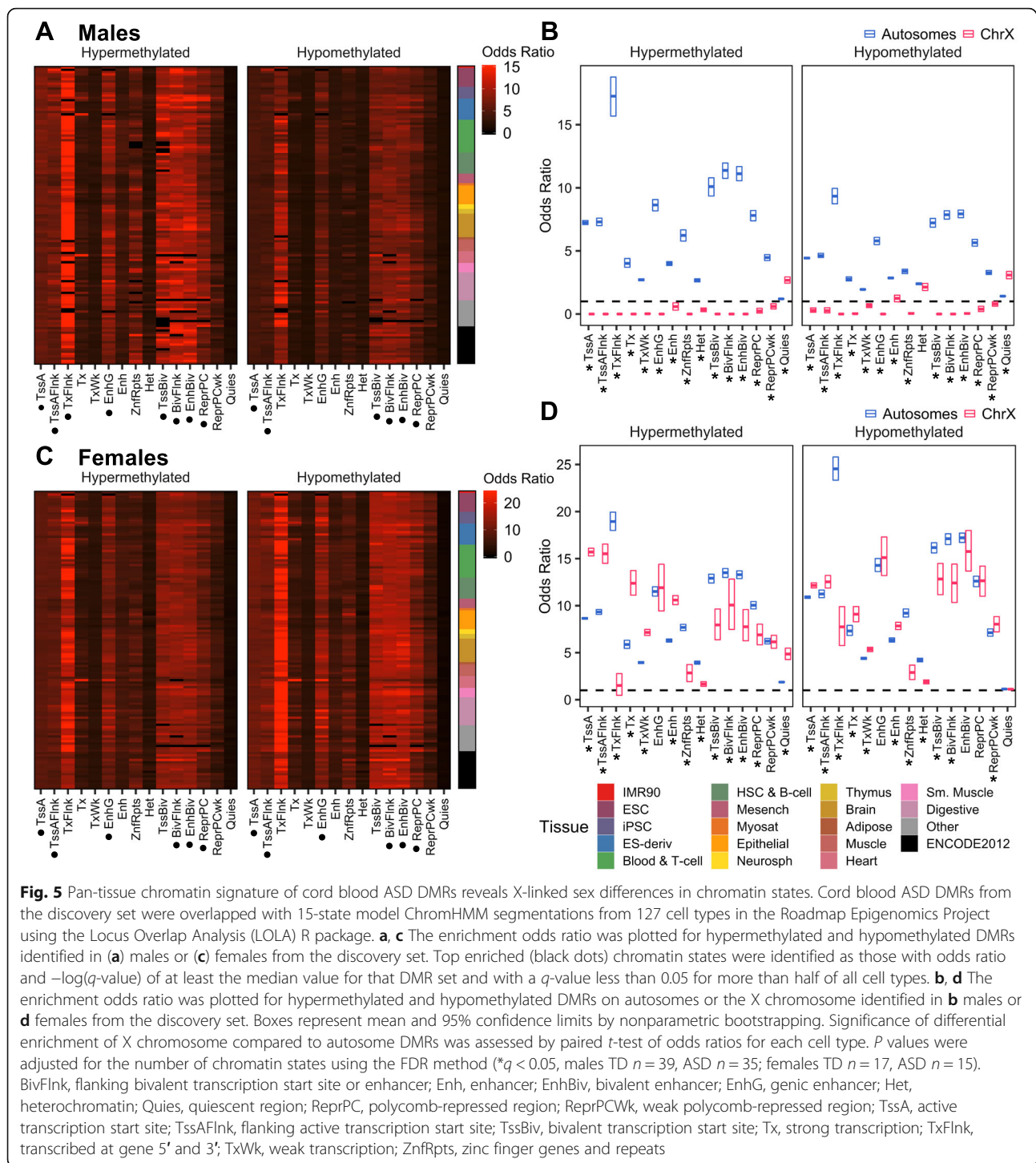
Fig. 4 Cord blood ASD DMR genes are significantly enriched for epigenetically dysregulated genes in ASD brain. Gene sets significantly enriched among ASD DMR genes in both discovery and replication sample sets for either males or females ($*q < 0.05$). Heatmaps show odds ratios for enrichment in genes annotated to DMRs relative to genes annotated to background calculated with Fisher's exact test for previously published studies of ASD and other neurological disorders. P values were adjusted using the false discovery rate (FDR) method for the total number of gene lists compared (pooled males TD $n = 56$, ASD $n = 56$; pooled females TD $n = 20$, ASD $n = 20$). Ac, acetylation; DEG, differentially expressed genes; Dup15, Chromosome 15q11-q13 Duplication syndrome; H3K27, histone 3 lysine 27; H3K4, histone 3 lysine 4; LCL, lymphoblastoid cell line; mCpG, CpG methylation; mCpH, CpH methylation; me3, trimethylation; PFC, prefrontal cortex; RTT, Rett syndrome; SFARI, Simons Foundation Autism Research Initiative

methodologies, no study examining differential gene expression in ASD subjects showed a significant overlap in gene hits with any of our ASD DMR genes from the discovery or replication set, including one conducted in cord blood from the MARBLES and EARLI cohorts [68] (Additional file 1: Fig. S21, Additional file 3: Table S29). The extensive overlap of cord blood DMR genes with those identified in other epigenomic studies across several tissues, combined with the lack of overlap with differentially expressed genes, supports the hypothesis that the DNA methylome is less time dependent than the transcriptome and more likely reflects past alterations that occurred in early development.

ASD DMRs are enriched for a pan-tissue epigenomic signature that differs between males and females on the X chromosome

To test the hypothesis that cord blood ASD DMRs were reflecting chromatin differences in *cis*-regulatory regions important in early prenatal life, we examined the enrichment of autosomal and X chromosomal DMRs for histone PTMs and chromatin states across many tissues and stages. Cord blood ASD DMRs were positionally compared with histone

PTM chromatin immunoprecipitation sequencing (ChIP-seq) peaks and ChromHMM-defined chromatin states in 127 cell types from the Roadmap Epigenomics Project [30]. ASD hyper- and hypomethylated DMRs in males and females in both sample sets were enriched for chromatin states involved in transcriptional regulation, including active transcription start sites (TssA and TssAFlnk) and bivalent enhancers (EnhBiv), across tissue types ($q < 0.05$; Fig. 5a,c; Additional file 1: Fig. S25A,C, Additional file 3: Table S31-S34). These chromatin states corresponded to significant enrichment in regions with H3K4me3, H3K27me3, and H3K4me1 across tissues ($q < 0.05$; Additional file 1: Fig. S26A,C, S27A,C, Additional file 3: Table S31-S34). In males, hypermethylated DMRs were enriched in genic enhancers (EnhG), while hypomethylated DMRs were enriched near bivalent (BivFlnk) and polycomb-repressed (ReprPC) regions in both sample sets ($q < 0.05$). In females, both hyper- and hypomethylated DMRs were enriched in EnhG, BivFlnk, and ReprPC regions in discovery and replication sets ($q < 0.05$). Together, these results demonstrate a sex-independent epigenetic signature of poised bivalent genes and enhancers that is pan-tissue, rather than limited to blood or immune-specific functions.



Interestingly, males and females displayed divergent patterns of ASD DMR enrichment in chromatin states and histone PTMs on the X chromosome versus autosomes. In males, hyper- and hypomethylated X chromosomal ASD DMRs in both sample sets were uniquely enriched for quiescent regions (Quies), and less enriched for all other chromatin states and corresponding histone

PTMs compared to autosomal DMRs ($q < 0.05$; Fig. 5b, Additional file 1: Fig. S25B, S26B, S27B, Additional file 3: Table S31, S33). In contrast to males, X-linked ASD DMRs in females were strikingly more enriched in active transcribed chromatin states (TssA, TssAFlnk, Tx, TxWk, and Enh) but less enriched in flanking transcribed regions (TxFlnk) and heterochromatin (Het) compared to

autosomal DMRs ($q < 0.05$; Fig. 5d, Additional file 1: Fig. S25D, Additional file 3: Table S32, S34). Hypermethylated ASD DMRs in females were also less enriched in other repressed or bivalent regions (ZnfRpts, TssBiv, BivFlnk, EnhBiv, and ReprRC) compared to autosomal DMRs ($q < 0.05$). These chromatin state differences in female X-linked DMRs corresponded with an increased enrichment for H3K4me1, H3K4me3, and H3K36me3 and a decreased enrichment for H3K9me3 ($q < 0.05$, Additional file 1: Fig. S26D, S27D). A CpG island analysis also confirmed the chromatin state enrichment differences between autosomal and X-linked ASD DMRs in males versus females (Additional file 1: Fig. S28, Additional file 3: Table S35). Together, these results suggest that in both males and females, autosomal ASD DMRs represent euchromatic, actively repressed, or bivalent chromatin states, while male-specific X-linked ASD DMRs reflect repressed quiescent heterochromatic regions and female-specific X-linked ASD DMRs encompass euchromatic, transcriptionally active regions.

ASD DMRs are enriched for binding motifs of methylation-sensitive transcription factors relevant to the fetal brain epigenome

Since DNA methylation may modify the actions of transcription factors, we hypothesized that specific methyl-sensitive transcription factor binding sites could be identified in ASD DMRs, which may give further insight into their functional relevance in early development and X-linked sex differences. To test this hypothesis, we examined enrichments of transcription factor binding site motifs within sex-specific ASD DMRs split by autosome versus X chromosome location. In both males and females, hypomethylated autosomal DMRs were enriched in motifs for HOXA2, TBX20, and ZNF675 ($q < 0.05$; Fig. 6a, Additional file 3: Table S36). However, for X-linked DMRs, none of the top enriched motifs were in common between males and females (Fig. 6b).

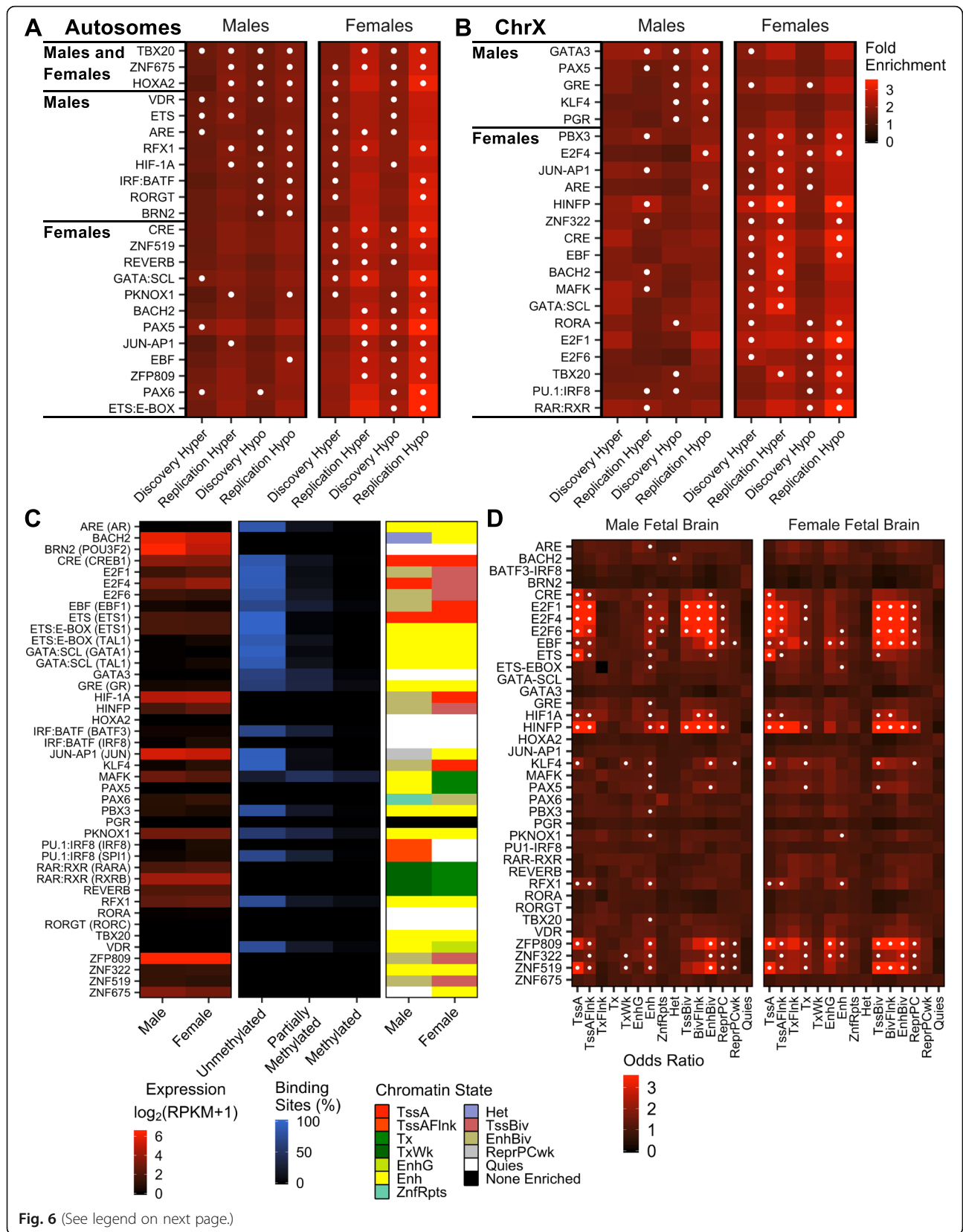
Among autosomal DMRs in males, hypermethylated DMRs were enriched for ETS and VDR motifs, while hypomethylated DMRs were also enriched for VDR as well as ARE, BRN2, HIF1A, IRF-BATF, RGX1, and RORGT motifs ($q < 0.05$; Fig. 6a). In contrast, hypomethylated X-linked DMRs in males were enriched for GATA3, GRE, KLF4, PAX5, and PGR motifs ($q < 0.05$; Fig. 6b). Unlike in males, three motifs were in common between female autosomal and X-linked DMRs (hyper: CRE and GATA:SCL; hypo: TBX20; Fig. 6a,b, Additional file 3: Table S36). Still, most of the top enriched motifs were unique for female autosomal versus X-linked ASD DMRs. Specifically, motifs for REVERB and ZNF519 were overrepresented among hypermethylated autosome DMRs, while ZNF519, and also BACH2, CRE, EBF, ETS:

E-BOX, JUN-AP1, PAX5, PAX6, PKNOX1, and ZFP809 motifs were overrepresented among hypomethylated autosome DMRs ($q < 0.05$; Fig. 6a). For X-linked DMRs in females, hypermethylated DMRs were enriched in ARE, BACH2, E2F4, EBF, HINFP, JUN-AP1, MAFK, PBX3, and ZNF322 motifs, while hypomethylated DMRs were enriched in PBX3 and E2F4, and also E2F1, E2F6, PU.1:IRF8, RAR:RXR, and RORA motifs ($q < 0.05$; Fig. 6b). The enrichment of largely distinct sets of motifs in ASD DMRs from males and females and from autosomes and the X chromosome suggests mostly separate groups of transcription factor motifs are associated with these DMRs.

To aid in the functional interpretation of these binding motifs in the context of methylation alterations in neurodevelopment, we investigated their transcription factor fetal brain expression, methylation sensitivity, and enrichment for fetal brain chromatin states. Transcription factors expressed in fetal brain with motifs enriched in male autosomal ASD DMRs were *BRN2/POU3F2*, *ETS1*, *HIF1A*, *RFX1*, and *ZNF675* (reads per kilobase per million reads (RPKM) > 1), while those enriched in male X-linked DMRs were not expressed in fetal brain (Fig. 6c). Of these expressed transcription factors, *ETS1* and *RFX1* prefer unmethylated motifs (less than 10% methylation). Transcription factors with female autosome DMR-enriched motifs that were expressed in fetal brain included *BACH2*, *CREB1*, *ETS1*, *JUN*, *PAX6*, *PKNOX1*, *REVERB*, *RFX1*, *ZNF519*, and *ZNF675* (RPKM > 1). Notably, *CREB1*, *ETS1*, *JUN*, *PKNOX1*, and *RFX1* bind selectively to unmethylated motifs. Among transcription factors with female X-linked DMR-enriched motifs, *BACH2*, *CREB1*, *E2F1*, *E2F4*, *E2F6*, *HINFP*, *JUN*, *MAFK*, *RARA*, *RXRB*, and *ZNF322* were expressed (RPKM > 1). Out of these expressed transcription factors, *CREB1*, *E2F1*, *E2F4*, *E2F6*, and *JUN* prefer unmethylated motifs, while *MAFK* prefers partially methylated motifs (10–90% methylation). ASD DMR-associated transcription factor motifs also significantly overlapped with chromatin states in fetal brain involved in transcriptional regulation, including transcription start sites (TssA, TssAFlnk, and TssBiv) and enhancers (Enh, EnhG, and EnhBiv; Fig. 6d, Additional file 3: Table S37). Together, these results support a functional role for differential methylation patterns identified in cord blood to impact transcription factors relevant to the fetal brain and highlight important sex differences.

Discussion

In this study, we found evidence that a DNA methylation signature of ASD relative to TD outcome exists in cord blood and that specific regions are consistently differentially methylated despite both technical and individual differences between data sets. Replication was



(See figure on previous page.)

Fig. 6 Cord blood ASD DMRs are significantly enriched for motifs for fetal brain-relevant methyl-sensitive transcription factors. Sequences in ASD DMRs were analyzed for enrichment in known transcription factor binding motifs compared to background regions using the Hypergeometric Optimization of Motif EnRichment (HOMER) tool. The fold enrichment for top enriched motifs was plotted for the indicated DMR sets for **a** DMRs on autosomes or **b** DMRs on chromosome X. Motifs were ordered by replicated sex and direction and plotted on the same scale. Top enriched (white dots) motifs were identified as those with a q -value less than 0.05 and present in the top quartile of fold enrichment and $-\log(q\text{-value})$ within that DMR set. Plotted motifs were the top enriched in both discovery and replication sets of DMRs for that sex and direction. **c** Expression of transcription factors with top enriched motifs in male or female fetal brain; proportion of ChIP-seq peaks in unmethylated (less than 10%), partially methylated (10–90%), and methylated (greater than 90%) contexts; and top ranked chromatin state by mean rank of odds ratio and p value for motif in male or female fetal brain and with a q -value less than 0.05. **d** Odds ratios for enrichment of top enriched motif locations in male or female fetal brain chromatin states. Top enriched (white dots) chromatin states were identified as those with a q -value less than 0.05 and present in the top quartile of odds ratio and $-\log(q\text{-value})$ within that tissue (pooled males TD $n = 56$, ASD $n = 56$; pooled females TD $n = 20$, ASD $n = 20$). RPKM, reads per kilobase per million reads

stronger at the level of broader gene regulatory domains and gene ontologies, rather than individual DMRs, although both were significantly higher than expected by chance. For instance, for broad gene regulatory domains defined by GREAT, 537 domains in males and 1762 domains in females were consistently associated with DMRs. In contrast, only 7 DMRs in males and 31 DMRs in females were regionally replicated in a locus-specific manner. However, replicated cord blood ASD DMR gene regulatory domains identified by our WGBS-based approach across two sequencing platforms significantly overlapped with genes identified from prior epigenetic ASD studies in brain and other tissues using array-based approaches. When a more conservative gene mapping approach was applied, discovery and replication ASD DMR genes still significantly overlapped with each other, with prior epigenetic ASD studies, and with X-linked and early neurodevelopmental functions. Together, these results demonstrate a convergence of common dysregulated genes and pathways and suggest that the methylation state of individual CpGs or specific CpG clusters is less replicable than the convergent epigenomic signature. In this way, the cord blood methylome ASD signature likely reflects the consequences of dysregulating functionally related sets of genes in early prenatal development, which is likely more important for the resulting phenotype than precisely altered CpGs.

Implicating their potential etiological relevance, cord blood ASD DMR gene regulatory domains were enriched for expression in brain, at the postsynaptic membrane, and during embryonic development in both males and females. Genes previously identified in epigenetic studies of ASD were also overrepresented among cord blood DMR genes. Of these, *RBFOX1* is particularly interesting, as it has been previously associated with ASD in post-mortem brain by studies of DNA methylation [22, 26], histone acetylation [27], and gene expression [13, 69, 70], as well as in genome-wide association studies [71]. *RBFOX1* encodes for a neuronal splicing factor, which was identified as the hub gene of an ASD-associated coexpression module in ASD post-mortem brain [100], and whose target genes are enriched among genes contributing to ASD risk [101]. Notably, *RBFOX1* was also found to be differentially methylated

in post-mortem brain of subjects with Rett syndrome and Dup15q syndrome, suggesting a fundamental role in neurodevelopment [24]. In contrast, cord blood ASD DMR genes were not enriched for those identified in transcriptome studies of ASD. This, together with the enrichment of cord DMR genes for embryonic expression, suggests the ASD-associated differential methylation identified in cord blood is more likely to be a remnant of past dysregulation in early development rather than a current correlate of transcript levels. The genes most likely to contribute to ASD are predicted to be expressed both in early prenatal development and in brain, which is indeed what we observe in our cord blood ASD DMR identified gene regulatory domains.

In a complementary approach to gene-level analyses, we investigated the region-based enrichment of ASD DMRs for CpG context, histone PTMs, chromatin states, and transcription factor motifs, and found similar epigenomic signatures for autosome DMRs in males and females. Overall, *cis*-regulatory regions present in almost all tissues were overrepresented among autosome DMRs, including TSSs, bivalent regions, and polycomb-repressed regions. Many transcription factors associated with ASD DMRs were specific to males or females, sensitive to methylation, and expressed in fetal brain. Specifically, *ETS1*, *RFX1*, and *CREB1* were enriched in specific fetal brain chromatin states, and all three displayed a sexually dimorphic pattern of enrichment in ASD DMRs, suggesting they may be involved in sex-specific transcriptional dysregulation in ASD. *ETS1* is widely expressed during embryogenesis and is involved in organ formation and morphogenesis of the mesoderm [102]. *RFX1* is necessary for the formation of the testis cord in the fetus and plays a role in spermatogenesis [103]. *CREB1* is important for neuronal stimulus-dependent gene expression, and its deletion in mice results in impaired survival of sensory and sympathetic neurons and axonal elongation [104]. Together, the functional implications of these findings are that cord blood ASD DMRs reflect early perturbations in neurodevelopment. Although ASD DMRs on autosomes are associated with similar chromatin states in both sexes, affected transcription factor binding sites are sex-specific.

Cord blood ASD DMRs identified from both sexes were enriched for X-linked gene regulatory domains and for distinctive chromatin states on the X chromosome compared to autosomes. Replicated ASD DMR genes on the X chromosome included 21 gene regulatory domains common to males and females, 34 only in males, and 152 only in females, with most of these genes previously associated with ASD in at least one genome-wide study. Compared to all genes in the genome, genes on the X chromosome are enriched for expression in brain and embryonic development and also mental retardation, suggesting that chromosome-wide epigenetic dysregulation of the X chromosome may predispose individuals to neurodevelopmental disorders. Additional evidence for this premise can be seen in X chromosome aneuploidy disorders such as Turner, Klinefelter, and trisomy X syndromes, where differences in neuroanatomy and behavior have been identified [105–110]. Interestingly, both Klinefelter and trisomy X subjects are at a more than fourfold increased risk of having ASD [110]. In addition, a large number of X-linked disorders involve intellectual disability [111].

X-linked ASD DMRs exhibited distinct chromatin and transcription factor features from those on autosomes, with notable sex differences. Male X-linked ASD DMRs were less enriched for active regulatory elements, but more enriched for quiescent regions compared to autosomal ASD DMRs. In contrast, female X-linked ASD DMRs were more enriched in transcriptionally active regions than autosomal ASD DMRs. Similar patterns were observed regardless of the direction of methylation change or tissue type. Additionally, only female X-linked ASD DMRs were enriched for the E2F1, E2F4, and E2F6 transcription factors, which are all expressed in fetal brain, are sensitive to methylation, and regulate cell cycle progression [112]. The disruption of active regulatory regions in females can also be seen in the large number of replicated X chromosome DMR genes in females compared to males, including *MECP2*, which was previously found to have altered methylation in female ASD brain [113]. Together, these results demonstrate that female-specific X-linked ASD DMRs predominate in euchromatic regions, while those specific to males are predominantly heterochromatic, suggesting sex differences due to epigenetic mechanisms on the X chromosome.

One potential explanation for this distinctive female pattern of epigenetic dysregulation could be the phenomenon of “X chromosome erosion,” recently identified in female pluripotent stem cells [114]. During the process of random X chromosome inactivation that occurs in human peri-implantation embryos, the long noncoding RNA *XIST* coats all X chromosomes in males and females, while another primate-specific X-linked long noncoding RNA named *XACT* coats only active X chromosomes to prevent transcriptional repression [115]. When X

chromosome inactivation becomes eroded in cultured human embryonic stem cells, *XACT* is aberrantly expressed, resulting in decreased *XIST*, loss of repressive H3K27me3 and DNA methylation at promoters, and transcriptional reactivation of some genes on the inactive X [114]. Notably, *XACT* was associated with ASD DMRs in both males and females, and its expression is normally exclusive to a preimplantation developmental window between 4 and 8 cells and the epiblast. We therefore hypothesize that aberrant *XACT* expression and/or X chromosome erosion during early development could underlie our findings of X-linked epigenetic dysregulation in cord blood from newborns later diagnosed with ASD. A second X chromosome in females may then serve as a mechanism of epigenetic protection against ASD and contribute to the observed 3 to 1 male bias in ASD [15].

A balanced interpretation of the results of this study should be guided by its strengths and limitations. The strengths of this study lie in its design and approach. Subjects were obtained from two prospective high-familial risk cohorts that were deeply phenotyped during the first 3 years of life, which included the gold standard ADOS assessment. The high-familial risk design increases ASD prevalence in a prospective cohort, as siblings of an ASD proband are significantly more likely to develop ASD or other developmental delays [116, 117]. Notably, this type of design also increases the power to detect environmental factors contributing to ASD risk [118]. In this instance, it is possible that the high-familial risk design could increase the ability to capture changes in DNA methylation related to either inherited early developmental programming or a marker of environmentally induced risk. In our analytical approach, we used WGBS to assay DNA methylation at more than 20 million CpGs and stratified subjects for DMR identification by both sex and sequencing platform, which were two major drivers of variability in methylation. Furthermore, we focused our analysis on systems-level features informed by chromosome biology, including gene locations and functions, as well as chromatin states and transcription factors at *cis*-regulatory regions. Findings were confirmed by replication in two groups of participants, different sequencing platforms, different bioinformatic approaches, and prior methylation studies of ASD in other tissues and platforms.

There are multiple limitations of this study, the first of which is that findings in these high-familial risk cohorts may not translate to low-risk populations. Second is the inherent limitation in statistical power because of the relatively small sample size due to the prospective study design and the uniqueness of the MARBLES and EARLI high-familial ASD risk cohorts. This study was not intended to identify individual DMRs with high confidence, but rather to test hypotheses about genome-level signatures of differential methylation in cord blood from

ASD subjects and their potential functional impact. A conservative power calculation estimated that this study was powered to detect DNA methylation differences of >15% between ASD and TD samples in all except the female replication group (Additional file 3: Table S38). Since methylation differences less than 15% are expected to be biologically informative, specific ASD DMRs identified in this study with small methylation differences would need to be validated, particularly in a larger study of 50–250 samples per group, based on these estimates. Additional DMRs with small methylation differences may also be identified by larger studies with greater power. Third, the conclusions about ASD DMR genes were primarily based on a relatively liberal approach to gene mapping based on broad regulatory domains and overlap to the closest gene, which may not be entirely accurate for some trans-acting regulatory elements. Fourth, cord blood is a complex tissue with dozens of different cell types that are undergoing dynamic changes in the prenatal period. Because we analyzed this tissue in bulk, the DMRs identified could reflect methylation alterations within individual cell types as well as changes in the proportions of these cells between subjects, potentially reducing power to detect disease relevant signatures. Notably, we identified an increased proportion of nRBCs in males with ASD, which correlated with lower global CpG methylation; however, nRBCs only correlated with methylation at a small number of ASD DMRs and those called with nRBC proportion as an adjustment covariate significantly overlapped with those without such adjustment. Furthermore, our demonstration of a lack of strong correlation of methylation within ASD DMRs with estimated cell proportions combined with the replication in ASD brain studies and tissue independence of the *cis*-regulatory regions identified alleviates these important concerns. Additionally, ASD is a heterogeneous disorder with high variability between subjects in genetic and environmental risk factors and comorbid phenotypes, which together with the small sample size and dichotomous outcome, may have limited our ability to detect small methylation differences in individual genes or CpG sites at high confidence. However, for those identified differences, we verified associations between methylation and continuous ASD symptom severity and cognitive functioning with more power and confirmed the lack of associations with non-neurodevelopmental factors. We supplemented this approach by identifying regions with permutation-based *p* value significance across two groups of subjects to reduce false positives and focused on convergent genes, gene pathways, and chromatin-level features. Lastly, we used a machine learning approach trained on the discovery samples to predict ASD versus TD outcome in the replication samples at an accuracy better than expected from random chance.

Conclusions

In this first study of the entire cord blood methylome from newborns later diagnosed with ASD, we identified differential methylation at birth at multiple levels, including regions, genes, functional gene sets, and chromatin states, and replicated these findings across two groups. In both males and females, ASD DMR-associated genes were more likely to be located on the X chromosome, to be expressed in brain and at the postsynaptic membrane, and to modify transcription factor binding sites relevant to the developing brain. Autosomal ASD DMRs were also enriched for *cis*-regulatory functional regions present across most tissues, including transcription start sites, CpG islands, enhancers, and bivalent regions. ASD DMRs on the X chromosome reflected chromosome- and sex-specific dysregulation, with an enrichment for quiescent regions among male DMRs, and an enrichment in open chromatin states in female DMRs, compared to those on autosomes. These findings in cord blood suggest that epigenetic dysregulation in ASD may originate during early prenatal development in a sex-specific manner and converge on brain-relevant genes to disrupt neurodevelopment.

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s13073-020-00785-8>.

Additional file 1: Figure S1. Global CpG methylation is associated with diagnosis and behavioral outcome scores only in males. **Figure S2.** The proportion of nRBCs is associated with behavioral outcome and global CpG methylation in males. **Figure S3.** Methylation at ASD DMRs in the discovery set is specifically associated with behavioral outcome. **Figure S4.** DMRs identified in all discovery subjects distinguish ASD from TD subjects. **Figure S5.** Sequencing platform has a larger effect on 10 kb window methylation than ASD diagnosis, sex, or study. **Figure S6.** DMRs identified in males and females distinguish ASD from TD subjects in the replication set. **Figure S7.** Methylation at ASD DMRs in the replication set is specifically associated with behavioral outcome. **Figure S8.** DMRs identified in all replication subjects distinguish ASD from TD subjects. **Figure S9.** The majority of ASD DMRs do not overlap probes on the 450 K and EPIC arrays. **Figure S10.** DMR methylation assayed by WGBS correlates with DMR methylation assayed by bisulfite pyrosequencing. **Figure S11.** A subset of ASD diagnosis DMRs are associated with ASD severity in independent sample sets. **Figure S12.** Machine learning methods workflow. **Figure S13.** Summary of machine learning datasets and results for males. **Figure S14.** Summary of machine learning datasets and results for females. **Figure S15.** ASD DMRs identified with adjustment for sex miss many genes found when stratifying for sex. **Figure S16.** Replicated ASD DMB genes overlap with replicated ASD DMR genes. **Figure S17.** ASD DMB genes are enriched for membrane, cell adhesion, and embryo-expressed genes. **Figure S18.** Neurodevelopmental genes are overrepresented on the X chromosome. **Figure S19.** Replicated DMR genes on the X chromosome are expressed in fetal brain. **Figure S20.** Female-specific replicated DMR genes on the X chromosome are expressed in fetal brain. **Figure S21.** Cord blood ASD DMR genes are significantly enriched for epigenetically dysregulated genes in ASD brain. **Figure S22.** Selected regions with replicated sex-independent DMR genes on the X chromosome. **Figure S23.** Selected regions with replicated male-specific DMR genes on the X chromosome. **Figure S24.** Selected regions with replicated female-specific DMR genes

on the X chromosome. **Figure S25.** ASD DMRs in replication subjects are differentially enriched for chromatin states on the X chromosome. **Figure S26.** ASD DMRs in discovery subjects are differentially enriched for histone PTMs on the X chromosome. **Figure S27.** ASD DMRs in replication subjects are differentially enriched for histone PTMs on the X chromosome. **Figure S28.** ASD DMRs on the X chromosome are enriched near CpG islands only in females.

Additional file 2: Table S1. Subject characteristics in relation to outcome at 36 months. **Table S2.** Subject characteristics table by subject. **Table S3.** Associations between ASD diagnosis, global CpG methylation, and the proportion of nRBCs. **Table S4.** ASD DMRs identified in discovery males with annotation and covariate association. **Table S5.** ASD DMRs identified in discovery females with annotation and covariate association. **Table S6.** ASD DMRs identified in males are maintained with adjustment for nRBC proportion. **Table S7.** ASD DMR methylation is associated with diagnosis after adjustment for all cord blood cell types. **Table S8.** ASD DMRs adjusting for sex identified in all discovery subjects with annotation and covariate association. **Table S9.** ASD DMRs identified in replication males with annotation and covariate association. **Table S10.** ASD DMRs identified in replication females with annotation and covariate association. **Table S11.** ASD DMRs adjusting for sex identified in all replication subjects with annotation and covariate association. **Table S12.** ASD DMRs replicate by region overlap with an independent computational method. **Table S13.** DMR methylation assayed by WGBS correlates with DMR methylation assayed by bisulfite pyrosequencing. **Table S14.** ASD DMRs replicate by region overlap with an independent sample set. **Table S15.** ASD DMRs replicated by region overlap with an independent sample set. **Table S16.** ASD DMRs replicated by differential methylation in an independent sample set. **Table S17.** ASD DMRs associated with autism severity among ASD subjects in multiple sample sets. **Table S18.** Replicated DMR and DMB genes with annotation. **Table S19.** ASD DMRs replicate by gene overlap with conservative mapping. **Table S20.** ASD DMBs identified in discovery males with annotation. **Table S21.** ASD DMBs identified in discovery females with annotation. **Table S22.** ASD DMBs identified in replication males with annotation. **Table S23.** ASD DMBs identified in replication females with annotation. **Table S24.** ASD DMB genes replicate in independent sample sets and overlap with DMR genes.

Additional file 3: Table S25. DMB gene functional enrichment for all comparisons. **Table S26.** DMR gene functional enrichment for all comparisons. **Table S27.** DMR gene functional enrichment for all comparisons with conservative mapping. **Table S28.** X-linked gene functional enrichment. **Table S29.** DMR gene ASD study enrichment for all comparisons. **Table S30.** DMR gene ASD study enrichment for all comparisons with conservative mapping. **Table S31.** Histone PTM and chromatin state enrichment for all, autosomal, and X-linked discovery males DMRs. **Table S32.** Histone PTM and chromatin state enrichment for all, autosomal, and X-linked discovery females DMRs. **Table S33.** Histone PTM and chromatin state enrichment for all, autosomal, and X-linked replication males DMRs. **Table S34.** Histone PTM and chromatin state enrichment for all, autosomal, and X-linked replication females DMRs. **Table S35.** CpG context enrichment for all, autosomal, and X-linked DMRs from all comparisons. **Table S36.** Transcription factor motif enrichment for all, autosomal, and X-linked DMRs from all comparisons. **Table S37.** Fetal Brain ChromHMM enrichment for top transcription factor motifs. **Table S38.** Post hoc power analysis.

Abbreviations

Ac: Acetylation; ADOS: Autism Diagnostic Observation Schedule; ASD: Autism spectrum disorder; BivFlnk: Flanking bivalent transcription start site or enhancer; ChIP-seq: Chromatin immunoprecipitation sequencing; chr: Chromosome; DAVID: Database for Annotation, Visualization, and Integrated Discovery; DEG: Differentially expressed genes; DMB: Differentially methylated block; DMR: Differentially methylated region; Dup15: Chromosome 15q11-q13 Duplication syndrome; EARLI: Early Autism Risk Longitudinal Investigation; Enh: Enhancer; EnhBiv: Bivalent enhancer; EnhG: Genic enhancer; EWAS: Epigenome-wide association study; FDR: False discovery rate; GREAT: Genomic Regions Enrichment of Annotations Tool; H3K27: Histone 3 lysine 27; H3K36: Histone 3 lysine 36; H3K4: Histone 3 lysine 4; H3K9: Histone 3 lysine 9; Het: Heterochromatin; HOMER: Hypergeometric

Optimization of Motif EnRichment; hyper: Hypermethylated; hypo: Hypomethylated; LCL: Lymphoblastoid cell line; LOLA: Locus Overlap Analysis; MARBLES: Markers of Autism Risk in Babies - Learning Early Signs; mCpG: CpG methylation; mCpH: CpH methylation; me3: Trimethylation; MSEL: Mullen Scales of Early Learning; nRBC: Nucleated red blood cell; PFC: Prefrontal cortex; PTM: Post-translational modification; Quies: Quiescent region; ReprPC: Polycomb-repressed region; ReprPCWk: Weak polycomb-repressed region; RPKM: Reads per kilobase per million reads; RTT: Rett syndrome; SD: Standard deviation; SFARI: Simons Foundation Autism Research Initiative; TD: Typically developing; TSS: Transcription start site; TssA: Active transcription start site; TssAFlnk: Flanking active transcription start site; TssBiv: Bivalent transcription start site; Tx: Strong transcription; TxFlnk: Transcribed at gene 5' and 3'; TxWk: Weak transcription; WGBS: Whole-genome bisulfite sequencing; ZnfRpts: Zinc finger genes and repeats

Acknowledgements

We would like to sincerely thank the participants in the MARBLES and EARLI studies.

Authors' contributions

CEM was the lead author and contributed substantially to DNA extraction, library preparation, all bioinformatic data analysis, data visualization, interpretation of results, and writing of the manuscript. JMJ performed bisulfite pyrosequencing and provided substantial assistance with sample organization, DNA extraction, library preparation, and bioinformatic data analysis. BIL contributed extensively to sequence alignment, quality control, and WGBS data analysis. HH performed machine learning analysis. YZ, KWD, and KMB added critical advice for bioinformatic data analysis methods and interpretation. JIF contributed to sample handling and organization. HEV, KL, LAC, CJN, SO, IH, MDF, and RJS contributed to the study design and also subject acquisition, diagnosis, and characterization. JML and RJS conceived of the study and funded the whole-genome bisulfite sequencing. JML contributed substantially to data interpretation and revision of the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by NIH grants: K12HD051958, P01ES011269, P30ES023513, R01ES016443, R01ES017646, R01ES020392, R01ES025531, R01ES025574, R01ES029213, R01ES028089, R24ES028533, U54HD079125, 1UG3OD023365, and UH3OD023365; EPA STAR grant RD-83329201; Autism Speaks grant 5938; Canadian Institutes of Health Research (CIHR) postdoctoral fellowships: MFE-146824, BPF-162684; and the MIND Institute IDDRC P50HD103526.

Availability of data and materials

The datasets supporting the conclusions of this article are available in the Gene Expression Omnibus repository at accession number GSE140730 [119]. All code used in the analyses for this study is available on GitHub [51].

Ethics approval and consent to participate

This study was performed in accordance with the ethical principles for medical research involving human subjects in the Declaration of Helsinki. The University of California, 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 to participate was obtained from the parents.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of Medical Microbiology and Immunology, Genome Center, and MIND Institute, University of California, Davis, CA, USA. ²Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI,

USA. ³Wendy Klag Center for Autism and Developmental Disabilities, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA. ⁴A. J. Drexel Autism Institute, Drexel University, Philadelphia, PA, USA. ⁵Division of Research, Kaiser Permanente Northern California, Oakland, CA, USA. ⁶Department of Biobehavioral Health, College of Health and Human Development, Pennsylvania State University, University Park, PA, USA. ⁷Psychiatry and Behavioral Sciences and MIND Institute, University of California, Davis, CA, USA. ⁸Department of Public Health Sciences and MIND Institute, University of California, Davis, CA, USA.

Received: 5 June 2020 Accepted: 25 September 2020

Published online: 14 October 2020

References

- Baio J, Wiggins L, Christensen DL, Maenner MJ, Daniels J, Warren Z, et al. Prevalence of autism spectrum disorder among children aged 8 years - autism and developmental disabilities monitoring network, 11 sites, United States, 2014. *Morb Mortal Wkly Rep Surveill Summ*. 2018;67(6):1–23.
- Sharma SR, Gonda X, Tarazi FI. Autism spectrum disorder: classification, diagnosis and therapy. *Pharmacol Ther*. 2018;190:91–104.
- Gotham K, Pickles A, Lord C. Standardizing ADOS scores for a measure of severity in autism spectrum disorders. *J Autism Dev Disord*. 2009;39(5):693–705.
- Masi A, DeMayo MM, Glozier N, Guastella AJ. An overview of autism spectrum disorder, heterogeneity and treatment options. *Neurosci Bull*. 2017;33(2):183–93.
- Bourgeron T. From the genetic architecture to synaptic plasticity in autism spectrum disorder. *Nat Rev Neurosci*. 2015;16(9):551–63.
- Bai D, Yip BHK, Windham GC, Sourander A, Francis R, Yoffe R, et al. Association of genetic and environmental factors with autism in a 5-country cohort. *JAMA Psychiatry*. 2019;76(10):1035–43.
- Grove J, Ripke S, Als TD, Mattheisen M, Walters RK, Won H, et al. Identification of common genetic risk variants for autism spectrum disorder. *Nat Genet*. 2019;51(3):431–44.
- Tick B, Bolton P, Happé F, Rutter M, Rijdsdijk F. Heritability of autism spectrum disorders: a meta-analysis of twin studies. *J Child Psychol Psychiatry*. 2016;57(5):585–95.
- Lyall K, Croen L, Daniels J, Fallin MD, Ladd-Acosta C, Lee BK, et al. The changing epidemiology of autism spectrum disorders. *Annu Rev Public Health*. 2017;38:81–102.
- Brown AS, Susser ES, Lin SP, Neugebauer R, Gorman JM. Increased risk of affective disorders in males after second trimester prenatal exposure to the Dutch hunger winter of 1944–45. *Br J Psychiatry*. 1995;166(5):601–6.
- Hertz-Picciotto I, Schmidt RJ, Krakowiak P. Understanding environmental contributions to autism: causal concepts and the state of science. *Autism Res*. 2018;11(4):554–86.
- Stevens HE, Smith KM, Rash BG, Vaccarino FM. Neural stem cell regulation, fibroblast growth factors, and the developmental origins of neuropsychiatric disorders. *Front Neurosci*. 2010;4:59.
- 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.
- Loomes R, Hull L, Mandy WPL. What is the male-to-female ratio in autism spectrum disorder? A systematic review and meta-analysis. *J Am Acad Child Adolesc Psychiatry*. 2017;56(6):466–74.
- Schaafsma SM, Pfaff DW. Etiologies underlying sex differences in autism spectrum disorders. *Front Neuroendocrinol*. 2014;35(3):255–71.
- Palmer N, Beam A, Agniel D, Eran A, Manrai A, Spettell C, et al. Association of sex with recurrence of autism spectrum disorder among siblings. *JAMA Pediatr*. 2017;171(11):1107–12.
- Werling DM, Geschwind DH. Recurrence rates provide evidence for sex-differential, familial genetic liability for autism spectrum disorders in multiplex families and twins. *Mol Autism*. 2015;6:27.
- Campisi L, Imran N, Nazeer A, Skokauskas N, Azeem MW. Autism spectrum disorder. *Br Med Bull*. 2018;127(1):91–100.
- Brown CJ, Grealley JM. A stain upon the silence: genes escaping X inactivation. *Trends Genet*. 2003;19(8):432–8.
- Vogel Ciernia A, LaSalle J. The landscape of DNA methylation amid a perfect storm of autism aetiologies. *Nat Rev Neurosci*. 2016;17(7):411–23.
- LaSalle JM, Powell WT, Yasui DH. Epigenetic layers and players underlying neurodevelopment. *Trends Neurosci*. 2013;36(8):460–70.
- Nardone S, Sams DS, Reuveni E, Getselter D, Oron O, Karpuz M, et al. DNA methylation analysis of the autistic brain reveals multiple dysregulated biological pathways. *Transl Psychiatry*. 2014;4:e433.
- Nardone S, Sams DS, Zito A, Reuveni E, Elliott E. Dysregulation of cortical neuron DNA methylation profile in autism spectrum disorder. *Cereb Cortex*. 2017;27(12):5739–54.
- Vogel Ciernia A, Laufer BI, Hwang H, Dunaway KW, Mordaunt CE, Coulson RL, et al. Epigenomic convergence of neural-immune risk factors in neurodevelopmental disorder cortex. *Cereb Cortex*. 2020;30(2):640–55.
- Ladd-Acosta C, Hansen KD, Briem E, Fallin MD, Kaufmann WE, Feinberg AP. Common DNA methylation alterations in multiple brain regions in autism. *Mol Psychiatry*. 2014;19(8):862–71.
- Wong CCY, Smith RG, Hannon E, Ramaswami G, Parikshak NN, Assary E, et al. Genome-wide DNA methylation profiling identifies convergent molecular signatures associated with idiopathic and syndromic autism in post-mortem human brain tissue. *Hum Mol Genet*. 2019;28(13):2201–11.
- Sun W, Poschmann J, Cruz-Herrera Del Rosario R, Parikshak NN, Hajan HS, Kumar V, et al. Histone acetylome-wide association study of autism spectrum disorder. *Cell*. 2016;167(5):1385–97.
- Berko ER, Suzuki M, Beren F, Lemetre C, Alaimo CM, Calder RB, et al. Mosaic epigenetic dysregulation of ectodermal cells in autism spectrum disorder. *PLoS Genet*. 2014;10(5):e1004402.
- Feinberg JL, Bakulski KM, Jaffe AE, Tryggvadottir R, Brown SC, Goldman LR, et al. Paternal sperm DNA methylation associated with early signs of autism risk in an autism-enriched cohort. *Int J Epidemiol*. 2015;44(4):1199–210.
- Zhu Y, Mordaunt CE, Yasui DH, Marathe R, Coulson RL, Dunaway KW, et al. Placental DNA methylation levels at CYP2E1 and IRS2 are associated with child outcome in a prospective autism study. *Hum Mol Genet*. 2019;28(16):2659–74.
- Hannon E, Schendel D, Ladd-Acosta C, Grove J. iPSYCH-Broad ASD Group, Hansen CS, et al. Elevated polygenic burden for autism is associated with differential DNA methylation at birth. *Genome Med*. 2018;10(1):19.
- Andrews SV, Ellis SE, Bakulski KM, Sheppard B, Croen LA, Hertz-Picciotto I, et al. Cross-tissue integration of genetic and epigenetic data offers insight into autism spectrum disorder. *Nat Commun*. 2017;8(1):1011.
- Lappalainen T, Grealley JM. Associating cellular epigenetic models with human phenotypes. *Nat Rev Genet*. 2017;18(7):441–51.
- 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.
- 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.
- Lord C, Risi S, Lambrecht L, Cook EH, 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.
- 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.
- Mullen EM. Mullen scales of early learning. MN: AGS Circle Pines; 1995.
- 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):1317–1327.e1.
- 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.
- Benjamini Y, Yekutieli D. The control of the false discovery rate in multiple testing under dependency. *Ann Stat*. 2001;29(4):1165–88.
- Clark SJ, Harrison J, Paul CL, Frommer M. High sensitivity mapping of methylated cytosines. *Nucleic Acids Res*. 1994;22(15):2990–7.
- Laufer BI, Hwang H, Vogel Ciernia A, Mordaunt CE, LaSalle JM. Whole genome bisulfite sequencing of Down syndrome brain reveals regional DNA hypermethylation and novel disorder insights. *Epigenetics*. 2019;14(7):672–84.
- Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for bisulfite-seq applications. *Bioinformatics*. 2011;27(11):1571–2.
- Laufer BI. CpG_Me: A whole genome bisulfite sequencing (WGBS) pipeline for the alignment and QC of DNA methylation that goes from raw

- reads (FastQ) to a CpG count matrix (Bismark cytosine reports). GitHub. 2020. https://github.com/ben-laufer/CpG_Me. Accessed 1 Oct 2018.
46. Krueger F. TrimGalore: A wrapper around Cutadapt and FastQC to consistently apply adapter and quality trimming to FastQ files, with extra functionality for RRBS data. GitHub. 2020. <https://github.com/FelixKrueger/TrimGalore>. Accessed 1 Oct 2018.
 47. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9(4):357–9.
 48. Wingett SW, Andrews S. FastQ Screen: A tool for multi-genome mapping and quality control. *F1000Res*. 2018;7:1338.
 49. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25(16):2078–9.
 50. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*. 2016;32(19):3047–8.
 51. Mordaunt CE. AutismCordBloodMethylation: Analysis of DNA methylation data by whole-genome bisulfite sequencing in umbilical cord blood from ASD and TD subjects in the MARBLES and EARLI studies. GitHub. 2020. <https://github.com/cemordaunt/AutismCordBloodMethylation>. Accessed 7 May 2020.
 52. Vu V. ggbiplot: A biplot based on ggplot2. GitHub. 2015. <https://github.com/vqv/ggbiplot>. Accessed 19 Mar 2020.
 53. Bakulski KM, Feinberg JI, Andrews SV, Yang J, Brown S, L McKenney S, et al. DNA methylation of cord blood cell types: applications for mixed cell birth studies. *Epigenetics*. 2016;11(5):354–362.
 54. Arney MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, et al. Minfi: a flexible and comprehensive bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics*. 2014;30(10):1363–9.
 55. Hicks SC, Irizarry RA. methylCC: technology-independent estimation of cell type composition using differentially methylated regions. *Genome Biol*. 2019;20(1):261.
 56. Laufer BI. DMRichR: An executable and package for the statistical analysis and visualization of differentially methylated regions (DMRs) from CpG count matrices (Bismark cytosine reports). GitHub. 2020. <https://github.com/ben-laufer/DMRrichR>. Accessed 1 Nov 2018.
 57. Korthauer K, Chakraborty S, Benjamin Y, Irizarry RA. Detection and accurate false discovery rate control of differentially methylated regions from whole genome bisulfite sequencing. *Biostatistics*. 2019;20(3):367–83.
 58. Hansen KD, Langmead B, Irizarry RA. BSmooth: from whole genome bisulfite sequencing reads to differentially methylated regions. *Genome Biol*. 2012;13(10):R83.
 59. Michels KB, Binder AM, Dedeurwaerder S, Epstein CB, Grealley JM, Gut I, et al. Recommendations for the design and analysis of epigenome-wide association studies. *Nat Methods*. 2013;10(10):949–55.
 60. Josse J, Husson F. missMDA: a package for handling missing values in multivariate data analysis. *J Stat Softw*. 2016;70(1):1–31.
 61. Gel B, Díez-Villanueva A, Serra E, Buschbeck M, Peinado MA, Malinverni R. regioneR: an R/Bioconductor package for the association analysis of genomic regions based on permutation tests. *Bioinformatics*. 2016;32(2):289–91.
 62. McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, et al. GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol*. 2010;28(5):495–501.
 63. Shen L. GeneOverlap: test and visualize gene overlaps. Bioconductor. doi: <https://doi.org/10.18129/B9.bioc.GeneOverlap> (2020).
 64. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics*. 2012;28(6):882–3.
 65. Kuhn M. Building predictive models in R using the caret package. *J Stat Softw*. 2008;28(5):1–26.
 66. Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res*. 2009;37(1):1–13.
 67. Fresno C, Fernández EA. RDAVIDWebService: a versatile R interface to DAVI D. *Bioinformatics*. 2013;29(21):2810–1.
 68. Mordaunt CE, Park BY, Bakulski KM, Feinberg JI, Croen LA, Ladd-Acosta C, et al. 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. *Mol Autism*. 2019;10(1):36.
 69. Gandal MJ, Haney JR, Parikshak NN, Leppa V, Ramaswami G, Hartl C, et al. Shared molecular neuropathology across major psychiatric disorders parallels polygenic overlap. *Science*. 2018;359(6376):693–7.
 70. Gandal MJ, Zhang P, Hadjimihael E, Walker RL, Chen C, Liu S, et al. Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science*. 2018;362(6420):eaat8127.
 71. 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.
 72. Andrews SV, Sheppard B, Windham GC, Schieve LA, Schendel DE, Croen LA, et al. Case-control meta-analysis of blood DNA methylation and autism spectrum disorder. *Mol Autism*. 2018;9:40.
 73. Gillissen C, Hehir-Kwa JY, Thung DT, van de Vorst M, van Bon BWM, Willemsen MH, et al. Genome sequencing identifies major causes of severe intellectual disability. *Nature*. 2014;511(7509):344–7.
 74. Iossifov I, O’Roak BJ, Sanders SJ, Ronemus M, Krumm N, Levy D, et al. The contribution of de novo coding mutations to autism spectrum disorder. *Nature*. 2014;515(7526):216–21.
 75. Kochinke K, Zweier C, Nijhof B, Fencikova M, Cizek P, Honti F, et al. Systematic phenomics analysis deconvolutes genes mutated in intellectual disability into biologically coherent modules. *Am J Hum Genet*. 2016;98(1):149–64.
 76. Sanders SJ, He X, Willsey AJ, Ercan-Sencicek AG, Samocha KE, Cicek AE, et al. Insights into autism spectrum disorder genomic architecture and biology from 71 risk loci. *Neuron*. 2015;87(6):1215–33.
 77. 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.
 78. Wong CCY, Meaburn EL, Ronald A, Price TS, Jeffries AR, Schalkwyk LC, et al. Methyloomic analysis of monozygotic twins discordant for autism spectrum disorder and related behavioural traits. *Mol Psychiatry*. 2014;19(4):495–503.
 79. 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.
 80. Doan RN, Lim ET, De Rubeis S, Betancur C, Cutler DJ, Chiacchetti AG, et al. Recessive gene disruptions in autism spectrum disorder. *Nat Genet*. 2019;51(7):1092–8.
 81. Lin P, Nicholls L, Assareh H, Fang Z, Amos TG, Edwards RJ, et al. Transcriptome analysis of human brain tissue identifies reduced expression of complement complex C1Q genes in Rett syndrome. *BMC Genomics*. 2016;17:427.
 82. Lambert JC, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer’s disease. *Nat Genet*. 2013;45(12):1452–8.
 83. Nguyen A, Rauch TA, Pfeifer GP, Hu VW. Global methylation profiling of lymphoblastoid cell lines reveals epigenetic contributions to autism spectrum disorders and a novel autism candidate gene, RORA, whose protein product is reduced in autistic brain. *FASEB J*. 2010;24(8):3036–51.
 84. 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.
 85. Shulha HP, Cheung I, Whittle C, Wang J, Virgil D, Lin CL, et al. Epigenetic signatures of autism: trimethylated H3K4 landscapes in prefrontal neurons. *Arch Gen Psychiatry*. 2012;69(3):314–24.
 86. Ellis SE, Gupta S, Moes A, West AB, Arking DE. Exaggerated CpH methylation in the autism-affected brain. *Mol Autism*. 2017;8:6.
 87. Banerjee-Basu S, Packer A. SFARI gene: an evolving database for the autism research community. *Dis Model Mech*. 2010;3(3–4):133–5.
 88. Jirtle RL. Imprinted genes: by species. <http://geneimprint.org/site/genes-by-species>. Accessed 16 Aug 2019.
 89. Lunnon K, Smith R, Hannon E, De Jager PL, Srivastava G, Volta M, et al. Methyloomic profiling implicates cortical deregulation of ANK1 in Alzheimer’s disease. *Nat Neurosci*. 2014;17(9):1164–70.
 90. Lawrence M, Gentleman R, Carey V. rtracklayer: an R package for interfacing with genome browsers. *Bioinformatics*. 2009;25(14):1841–2.
 91. Sheffield NC, Bock C. LOLA: enrichment analysis for genomic region sets and regulatory elements in R and Bioconductor. *Bioinformatics*. 2016;32(4):587–9.
 92. Cavalcante RG, Sartor MA. annotatr: genomic regions in context. *Bioinformatics*. 2017;33(15):2381–3.

93. Consortium RE, Kundaje A, Meuleman W, Ernst J, Bilenyk M, Yen A, et al. Integrative analysis of 111 reference human epigenomes. *Nature*. 2015; 518(7539):317–30.
94. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell*. 2010;38(4):576–89.
95. Miller JA, Ding S-L, Sunkin SM, Smith KA, Ng L, Szafer A, et al. Transcriptional landscape of the prenatal human brain. *Nature*. 2014;508(7495):199–206.
96. Xuan Lin QX, Sian S, An O, Thieffry D, Jha S, Benoukrif T. MethMotif: an integrative cell specific database of transcription factor binding motifs coupled with DNA methylation profiles. *Nucleic Acids Res*. 2019;47(D1): D145–54.
97. Orr M, Liu P. Sample size estimation while controlling false discovery rate for microarray experiments using the *ssize.fdr* package. *R J*. 2009;1(1):47.
98. Burns TG, King TZ, Spencer KS. Mullen scales of early learning: the utility in assessing children diagnosed with autism spectrum disorders, cerebral palsy, and epilepsy. *Appl Neuropsychol Child*. 2013;2(1):33–42.
99. Melnyk S, Fuchs GJ, Schulz E, Lopez M, Kahler SG, Fussell JJ, et al. Metabolic imbalance associated with methylation dysregulation and oxidative damage in children with autism. *J Autism Dev Disord*. 2012;42(3):367–77.
100. 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.
101. De Rubeis S, He X, Goldberg AP, Poultnery CS, Samocha K, Cicek AE, et al. Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature*. 2014;515(7526):209–15.
102. Kola I, Brookes S, Green AR, Garber R, Tymms M, Papas TS, et al. The Ets1 transcription factor is widely expressed during murine embryo development and is associated with mesodermal cells involved in morphogenetic processes such as organ formation. *Proc Natl Acad Sci U S A*. 1993;90(16): 7588–92.
103. Wang B, Qi T, Chen S-Q, Ye L, Huang Z-S, Li H. RFX1 maintains testis cord integrity by regulating the expression of *Itga6* in male mouse embryos. *Mol Reprod Dev*. 2016;83(7):606–14.
104. Lonze BE, Riccio A, Cohen S, Ginty DD. Apoptosis, axonal growth defects, and degeneration of peripheral neurons in mice lacking CREB. *Neuron*. 2002;34(3):371–85.
105. Lin A, Clasen L, Lee NR, Wallace GL, Lalonde F, Blumenthal J, et al. Mapping the stability of human brain asymmetry across five sex-chromosome aneuploidies. *J Neurosci*. 2015;35(1):140–5.
106. Hong DS, Hoefl F, Marzelli MJ, Lepage J-F, Roeltgen D, Ross J, et al. Influence of the X-chromosome on neuroanatomy: evidence from Turner and Klinefelter syndromes. *J Neurosci*. 2014;34(10):3509–16.
107. Hong D, Scaletta Kent J, Kesler S. Cognitive profile of Turner syndrome. *Dev Disabil Res Rev*. 2009;15(4):270–8.
108. Lawrence K, Kuntsi J, Coleman M, Campbell R, Skuse D. Face and emotion recognition deficits in Turner syndrome: a possible role for X-linked genes in amygdala development. *Neuropsychology*. 2003;17(1):39–49.
109. Skakkebaek A, Moore PJ, Pedersen AD, Bojesen A, Kristensen MK, Fedder J, et al. Anxiety and depression in Klinefelter syndrome: the impact of personality and social engagement. *PLoS One*. 2018;13(11):e0206932.
110. Wilson AC, King J, Bishop DVM. Autism and social anxiety in children with sex chromosome trisomies: an observational study. *Wellcome Open Res*. 2019;4:32.
111. Ropers H-H, Hamel BCJ. X-linked mental retardation. *Nat Rev Genet*. 2005; 6(1):46–57.
112. Thurlings I, de Bruin A. E2F transcription factors control the roller coaster ride of cell cycle gene expression. *Methods Mol Biol*. 2016;1342:71–88.
113. Nagarajan RP, Patzel KA, Martin M, Yasui DH, Swanberg SE, Hertz-Picciotto I, et al. MECP2 promoter methylation and X chromosome inactivation in autism. *Autism Res*. 2008;1(3):169–78.
114. Vallot C, Quimette J-F, Makhlouf M, Féraud O, Pontis J, Côme J, et al. Erosion of X chromosome inactivation in human pluripotent cells initiates with XACT coating and depends on a specific heterochromatin landscape. *Cell Stem Cell*. 2015;16(5):533–46.
115. Vallot C, Patrat C, Collier AJ, Huret C, Casanova M, Liyakat Ali TM, et al. XACT noncoding RNA competes with XIST in the control of X chromosome activity during human early development. *Cell Stem Cell*. 2017;20(1):102–11.
116. 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.
117. Charman T, Young GS, Brian J, Carter A, Carver LJ, Chawarska K, et al. Non-ASD outcomes at 36 months in siblings at familial risk for autism spectrum disorder (ASD): a Baby Siblings Research Consortium (BSRC) study. *Autism Res*. 2017;10(1):169–78.
118. Weinberg CR, Shore DL, Umbach DM, Sandler DP. Using risk-based sampling to enrich cohorts for endpoints, genes, and exposures. *Am J Epidemiol*. 2007;166(4):447–55.
119. Mordaunt CE, Jianu JM, Laufer BI, Zhu Y, Hwang H, Dunaway KW, et al. Cord blood DNA methylome in newborns later diagnosed with autism spectrum disorder reflects early dysregulation of neurodevelopmental and X-linked genes datasets. *Gene Expression Omnibus*. 2020. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE140730>. Accessed 20 Nov 2019.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

