Impact of exposures to persistent endocrine disrupting compounds on the sperm methylome in regions associated with neurodevelopmental disorders

Angela G. Maggio
The George Washington University School of Medicine and Health Sciences
Henry T. Shu
The George Washington University School of Medicine and Health Sciences
Benjamin I. Laufer
University of California Davis
Hyeyeon Hwang
University of California Davis
Chongfeng Bi
The George Washington University School of Medicine and Health Sciences
Yinglei Lai
The George Washington University
Janine M. LaSalle
University of California Davis
Valerie W. Hu (valhu@gwu.edu)
George Washington University https://orcid.org/0000-0002-3357-0777

Research

Keywords: DNA methylation, sperm, Faroe Islands, endocrine disruptors, autism

Posted Date: March 17th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-297668/v1

Impact of exposures to persistent endocrine disrupting compounds on the sperm methylome in regions associated with neurodevelopmental disorders

Angela G. Maggio[1], Henry T. Shu[1,2], Benjamin I. Laufer[3], Hyeyeon Hwang[3], Chongfeng Bi[1], Yinglei Lai[4], Janine M. LaSalle[3], and Valerie W. Hu[1]

[1]Dept. of Biochemistry and Molecular Medicine; The George Washington University School of Medicine and Health Sciences; Washington, DC 20037 USA

[2] The Johns Hopkins University, School of Medicine; 733 N Broadway St, Baltimore, MD 21205 USA

[3] Medical Microbiology and Immunology, Genome Center, MIND Institute, Perinatal Origins of Disparities Center, and Environmental Health Sciences Center, UC Davis School of Medicine, Davis, CA 95616 USA

[4]Dept. of Statistics, The George Washington University, Washington, DC 20052 USA

Email addresses - Angela Maggio: agmaggio@gmail.com; Henry Shu: hshu5@jhmi.edu; Benjamin Laufer: blaufer@ucdavis.edu; Hyeyeon Hwang: hyehwang@ucdavis.edu; Chongfeng Bi: cbi55@gwu.edu; Yinglei Lai: ylai@gwu.edu; Janine LaSalle: jmlasalle@ucdavis.edu; Valerie Hu (corresponding author): valhu@gwu.edu, Department of Biochemistry and Molecular Medicine, The George Washington University School of Medicine and Health Sciences, 2300 Eye St., NW, Washington, DC 20037, USA.

Keywords: DNA methylation, sperm, Faroe Islands, endocrine disruptors, autism

Abstract

Background: Although autism spectrum disorder (ASD) is among the most heritable of neurodevelopmental disorders, the rapidly rising prevalence of ASD suggests that environmental factors may contribute to epigenetic modifications that can influence risk for ASD. Endocrine disrupting compounds (EDCs), such as the long-lived organochlorines, are of particular interest with respect to risk for autism because of their ability to interfere with sex hormones that have been implicated in ASD.

Methods: Whole genome bisulfite sequencing was used to identify genome-wide differentially methylated regions (DMRs) in a total of 52 sperm samples from a cohort of Faroese men who were equally divided into high and low exposure groups based on their serum levels of the long-lived organochlorine 1,1-dichloro-2,2-bis(p-chlorophenyl)ethylene (DDE). Inhabitants of the Faroe Islands, who as a group are considered a genetic isolate, may be exposed to higher than average levels of such persistent EDCs because of their native diet that is enriched in pilot whale meat and blubber. Gene ontology and pathway analyses were used to determine enrichment in ASD-relevant pathways and functions among the DMR-associated genes.

Results: DMRs were enriched in autism risk genes and could also discriminate between samples in high and low exposure groups. Functional and pathway analyses of these DMR-associated genes show significant enrichment for neurodevelopmental processes frequently impacted by ASD. Of note, the DMR-associated genes significantly overlap with those previously identified in sperm from fathers of children with ASD versus fathers of neurotypical children, thus suggesting a potential environmental mechanism for introducing ASD-associated epigenetic changes in the sperm methylome.

Limitations: A limitation of this study is the relatively low number of samples, although this is somewhat offset by the use of a genetic isolate which reduces the genetic heterogeneity that is

52 often a major challenge in epigenetic studies. Another limitation is the lack of a completely
53 unexposed set of samples for comparisons since persistent EDCs can be detected in a majority of
54 individuals.
55 Conclusion: Results of this study show that elevated exposure to certain organochlorines is
56 associated with genome-wide DNA methylation patterns in sperm affecting genes involved in
57 neurological functions and developmental disorders, including ASD.

Introduction

Autism spectrum disorder (ASD) is a highly complex neurodevelopmental disorder characterized by deficits in social communication, repetitive behaviors, and restricted interests (1). In a span of only 12 years from 2004 to 2016, the United States' Center for Disease Control and Prevention reported an increase in prevalence of ASD from roughly 1 in 125 children to 1 in 54, with a male to female sex bias of 4.3 to 1 (2). While increased awareness of the disorder and expansion of the diagnosis to include the moderate and milder forms of ASD (such as pervasive developmental disorder-not otherwise specified and Asperger syndrome) has been suggested to account for the increase in prevalence, epidemiological studies suggest that there may be environmental contributions to ASD as well (3). Similarly, although ASD exhibits a strong genetic component as revealed by concordance rates in studies on affected monozygotic versus dizygotic twins (4,5), the incomplete penetrance of ASD even in monozygotic twins also suggests a role for gene by environment (G x E) interactions which are often mediated through epigenetic mechanisms.

Early evidence for the involvement of epigenetics in neurodevelopmental disorders comes from studies on syndromic forms of ASD (e.g., Fragile X, Angelman's, Prader-Willi, and Rett syndromes) in which the etiologically relevant genes are either aberrantly methylated/imprinted or are involved in the recognition of methylated DNA sequences (6,7) . In 2010, we demonstrated for the first time genome-wide DNA methylation differences associated with idiopathic autism in lymphoblastoid cell lines (LCLs) from monozygotic twins and sib pairs discordant for diagnosis of autism (8). Importantly, by integrating these large-scale methylation analyses with genome-wide transcriptome analyses of the same cohort, we showed that the majority of the methylation differences were inversely associated with differences in gene

expression (9,10). Moreover, two of the validated hypermethylated genes, *BCL2* and *RORA*, were also found to be reduced in brain tissues from individuals with ASD (8), thus linking these molecular changes observed in blood-derived cells to autism brain pathology. We further showed that *RORA* could be regulated in opposite directions by male and female sex hormones, with male hormones suppressing expression (11). In addition, we demonstrated that over 2000 genes were putative transcriptional targets of this nuclear hormone receptor, including hundreds of autism risk genes (12). These findings led us to postulate that this hormonal dependence of *RORA* expression might contribute to the observed sex bias in ASD that has been hypothesized to be related to elevated fetal testosterone or steroidogenic activity (13,14). Moreover, we postulated that endocrine disrupting compounds (EDCs) that can mimic or interfere with normal hormonal signaling and metabolism might perturb the expression and function of this critical candidate gene for ASD, thereby increasing risk for autism (15).

Indeed, environmental exposures to EDCs have been linked with various diseases including neurodevelopmental disorders (16-19). Both experimental and epidemiological studies have reported harmful effects of exposures to both short- and long-lived EDCs, such as bisphenol A (BPA) and phthalates, and persistent organic pollutants (POPs) like polychlorinated biphenyl (PCB) and polybrominated diphenyl ether (PBDE) compounds, on neurodevelopment. Rodent studies have shown associations between EDC exposure and disruption of social hormones, social recognition, locomotion, and excitatory-inhibitory synapse pathways (20-22). Exposure to 3,3'-dichlorobiphenyl exerted effects on axonal and dendritic growth in primary rat neurons (23). Epidemiological studies further show that prenatal human exposures to EDCs, specifically DDE and PCBs, has been linked to alterations in hormone levels in offspring with potential for lasting and widespread signaling disruption (16,24).

105 With respect to mechanisms of action, many studies have investigated epigenetics as a
106 potential mediator of EDC-induced changes in phenotype and disease state. For example,
107 Manikkam et al. showed that certain plastics-derived EDCs induced transgenerational
108 inheritance of obesity, reproductive disease and epimutations in sperm (25). More recently,
109 McBirney et al. reported that exposure of pregnant rats during gestation to the herbicide atrazine
110 increased the frequency of testis disease and mammary tumors as well as induced changes in
111 body weight and onset of puberty in the F2 and F3 offspring (26). Additionally, they reported
112 DNA methylation changes in sperm, which were detected in offspring in each generation (26).
113 Regarding EDC involvement in autism, Dunaway et al. demonstrated that treatment of a
114 neuronal cell model with PCB 95 was associated with significant global DNA hypomethylation
115 of autism candidate genes with a direct impact on gene expression (27). In the same study,
116 hypomethylation was also observed in brain tissues from individuals with Dup15q syndrome (a
117 condition frequently associated with ASD) which had been previously associated with exposure
118 to PCB 95 (28). The role of epigenetics as an underlying mechanism for EDC-induced
119 neurodevelopmental disorders and ASD is further discussed in a number of recent reviews (29-
120 32).
121 In addition to having an impact on neurodevelopment, EDCs have also been implicated in
122 many other diseases or human conditions, such as cancer, diabetes, obesity, metabolic disorders,
123 and infertility (33). With respect to infertility, human exposures to a variety of organochlorines
124 have been shown to affect chromatin integrity (34) and sex chromosome aneuploidy in sperm
125 (35). EDC-associated sex chromosome aneuploidy in sperm was further confirmed in an
126 additional study involving a cohort of men in the Faroe Islands who have higher than average
127 exposures to POPs based on their diet which is rich in pilot whale meat and blubber (36). These

chromosomal changes in sperm suggested that EDCs may also induce epigenetic changes in sperm DNA.

In recent years, the sperm methylome (i.e., genome-wide DNA methylation pattern) has come under increasing scrutiny as a reservoir of epigenetic changes that may have an impact on a wide variety of human disorders, from infertility to neurodevelopmental and psychiatric disorders (37). In particular, DNA methylation changes in sperm have been reported as a function of advanced paternal age, in both mice and men (38,39). Notably, the changes have included genes that overlap with those reportedly associated with autism, schizophrenia, and bipolar disorder. Since advanced paternal age has also been associated with risk for autism (40-42), Feinberg et al. recently investigated DNA methylation differences in sperm from fathers with a child with ASD and from fathers of neurotypical children (43). They identified 193 DMRs in paternal sperm of ASD fathers, many of which were in the proximity of genes involved in developmental processes. A more recent study on the sperm methylome of fathers that have children with or without ASD further shows the potential for development of a biomarker screen for ASD based on DMRs (44). These studies thus establish a link between DNA methylation changes in sperm and possible risk for changes in sperm of the fathers with children affected by ASD.

Materials and methods

Aims, design, and setting of the study

The primary objectives of this study are to: 1) determine if high versus low exposures to the persistent organochlorine 1,1-dichloro-2,2-bis(p-chlorophenyl)ethylene (DDE) are associated with differentially methylated regions (DMRs) in sperm from a Faroese cohort; 2) determine if

genes associated with DDE DMRs are enriched for ASD risk genes; 3) identify pathways and functions among DMR-associated genes. Here, we used whole genome bisulfite sequencing (WGBS) to investigate differences in DNA methylation in sperm from 52 men exhibiting the highest (n = 26) and lowest (n = 26) exposures to DDE, a stable breakdown product of the insecticide DDT (dichloro-diphenyl-trichloroethane). The samples were divided into discovery (n = 32) and validation sets (n = 20) for the WGBS analyses. Gene ontology and pathway analyses were employed to determine enrichment in pathways and functions among the DMR-associated genes. Differential methylation of selected genes relevant to ASD and other neurodevelopmental disorders was further validated by pyrosequencing analyses.

Semen samples and demographics of donors

Raw ejaculate (i.e., semen) from 52 young men in a general population cohort of the Faroe Islands (Denmark) was kindly provided by Dr. Pál Weihe, Director of the Department of Occupational Medicine and Public Health of the Faroese Hospital System in Tórshavn, Faroe Islands. Inhabitants of the Faroe Islands are considered genetic isolates and therefore are expected to have similar genetic background and polymorphisms that would otherwise be considered major confounding variables in genomics studies. Samples were collected throughout 2007 and 2008 and de-identified by the Faroese Hospital System. Information related to height, weight, age, smoking status, length of abstinence, sample collection date, serum concentrations of the four most prevalent PCB congeners, (PCB 118, PCB 138, PCB 153, PCB 180) and DDE, were recorded for each sample as part of the Faroese General Population research cohort (which was part of the parent study approved by the local Science Ethics Committee for the Faroe Islands). The de-identified semen samples were made available for this study through a Data Processor Agreement between Dr. Pál Weihe at the Faroese Hospital System and Dr. Valerie Hu

at The George Washington University, together with the serum levels of the organochlorines that had been determined for each donor as previously described (45). A summary of demographic information and serum concentrations of EDCs for the cohort used in this study is provided in **Additional file 1**. Regression analyses of the potential covariates were performed using statistical packages in Excel. These analyses showed that while there were significant correlations between the concentrations of DDE with those of DDT and the sum of the four most prevalent PCB congeners (**Additional file 2**), there were no correlations between DDE levels and body mass index (BMI), smoking status, or any of the sperm parameters, including sperm concentration and mobility (**Additional files 3 and 4**). Faroese semen samples with a minimum sperm count of 20 x 10^6 per ml were divided by Dr. Melissa Perry (GWU) into three exposure groups based on the recorded levels of DDE in serum, and 26 samples each from the first (lowest exposure) and third (highest exposure) tertiles were provided for this methylation study. It should be noted that, once the 52 semen samples were received, they were processed through bisulfite sequencing analyses without knowledge of the exposure tertiles to which they were assigned.

Sperm isolation and DNA extraction methods

Sperm was isolated from semen (which contains various cell types) using a published discontinuous gradient protocol (46). Briefly, 100 µl of semen was washed with 2 mL Quinn's Sperm Washing Medium (Origio, Trumbull, CT), and cells were pelleted at 600 g for 5 minutes at 4°C. The pelleted cells were resuspended in 0.5 ml Quinn's solution and counted in a hemocytometer to determine the initial number of sperm and somatic cells in the semen sample. Next, a discontinuous gradient of PureCeption solution (Origio) diluted with Quinn's was formed in a 15 ml conical tube with 1.5 ml of 40% PureCeption over 1.5 ml of 90% PureCeption. The washed cells were layered on top of the gradient before centrifugation at 300 g

for 20 minutes at room temperature. The pellet was transferred to a new 15 ml tube and resuspended in 3 ml Quinn's Washing solution. The cells were pelleted at 600 g for 5 minutes, and then resuspended in 500 µl Quinn's after removal of the supernatant. Cells were recounted to determine total number of sperm and somatic cell contamination (which never exceeded 1%) before centrifugation at 4000 g for 1 minute.

DNA was isolated from the purified sperm cells using a Qiagen AllPrep DNA/RNA mini kit following the manufacturer's protocol. Sperm pellets were first lysed in 450 µl RLT buffer with 50 µl added tris(2-carboxyethyl)phosphine solution (TCEP, a bond-breaker) by vortexing with 0.1 gm RNase/DNase-free stainless steel microbeads for 5 min at RT in a Disruptor Genie (Scientific Industries, Bohemia, NY). Lysates were immediately stored at -80°C until DNA extraction which was usually completed the next day using the standard protocol.

Whole Genome Bisulfite Sequencing (WGBS)

Samples for WGBS were divided into two batches, a discovery set of 32 samples and a validation set of 20 samples. The samples were sent to Admera Health (South Plainfield, NJ), a CLIA-certified laboratory, for WGBS analyses. The directional Illumina TruSeq DNA Methylation library kit was used for sample preparation. WGBS was then performed on an Illumina HiSeq X sequencer resulting in 150 bp PE reads achieving roughly 4x coverage genome wide per sample. Raw FASTQ files were received from Admera Health for further analyses. As mentioned previously, all samples were processed blindly without the processors' knowledge of tertile level from sperm isolation through the WGBS analyses to minimize handling bias.

WGBS bioinformatics pipeline

We utilized a bioinformatics pipeline comprised of CpG_Me for alignment and DMRichR for differential methylation determination as published on github

(https://github.com/ben-laufer). CpG_Me builds on previously published bioinformatic tools and pipelines (47,48). The DMRichR workflow similarly builds on previously established bioinformatic packages such as dmrseq and bsseq (48-50). All WGBS data was analyzed on The George Washington University's high-performance cluster, Colonial One. First, raw read files (FASTA) were trimmed and quality checked with Trim-Galore and FASTQC software. Forward reads were trimmed by 8bp on the 3'- and 5'-ends. Reverse reads were trimmed 8bp on the, 3'-ends and 20bp on the 5'-ends in order to remove methylation bias often seen at the ends of reads. M-Bias plots from FASTQC analyses were examined to determine if trimming was sufficient. Trim Galore was also used to filter out bases with Phred scores lower than 20 that would indicate a 1 in 100 probability of an incorrect base call. Reads were then aligned to a reference human genome (hg38) using Bismark, a three-letter aligner. Methylation calls are differentiated among CpG, CHG, and CHH contexts (47), but only the CpG sites were considered in this study. Approximately 74% of bisulfite reads aligned to a bisulfite-converted reference human genome allowing for the assay of 10.08 million CpG sites. All raw and processed data from the WGBS analyses have been deposited into the NBCI's Gene Expression Omnibus (GEO) repository (GEO Accession number GSE165915).

Identification of differentially methylated regions (DMRs)

DMRs were identified using dmrseq and bsseq Bioconductor packages in the wrapped pipeline of DMRichR. Default parameters for the DMRichR executable script were used which included coverage set to 1x, per Group set to 100%, minCpGs set to 5, and maxPerms set to 10. Covariables adjusted for in DMR analysis included BMI, age, days in storage, batch effects of processor and date of processing, percent motile sperm, and smoking status. DMRs with permutation $p \leq 0.05$ were identified for both discovery and validation sample sets. Differences

247 in percent methylation (a measure of effect size) ranged from ±7 to ±38 for all DMRs, as shown
248 in **Additional files 5 and 6**. With DMRichR, DMRs were annotated for genomic and CG
249 context. Annotation resources within DMRichR include the annotatr and rGREAT (Genomic
250 Regions Enrichment of Annotations Tool) open source packages in Bioconductor. Annotatr (51)
251 was used to visualize and compare annotated genomic sites/regions (e.g., promoters, 5'UTRs,
252 exons, introns) that were identified within the discovery and validation sets, while rGREAT (52)
253 was used for mapping genes to the sites. The mapped genes were then utilized for pathway and
254 functional analyses using Ingenuity Pathway Analysis software (Qiagen, Germantown, MD) as
255 described below.
256 ***Gene ontology and network prediction analysis***
257 Gene ontology analysis of the DMRs assigned to genes was accomplished using two
258 different programs; GofuncR (53), which was modified for use with DMRichR, and the open-
259 access STRING Version 11.0 (Search Tool for the Retrieval of Interacting Genes/Proteins) (54).
260 GofuncR maps a DMR to a gene if it is between 5 kb upstream and 1 kb downstream of the gene
261 body and also uses information from the background regions. A custom meta p-value analysis
262 was performed using the sum of logs of the p-values of GO terms to integrate the data from the
263 discovery and validation sets, and then terms with a meta p-value < 0.05 were slimmed using
264 REVIGO (55) to reduce redundancy among the most significant GO terms. Ingenuity Pathway
265 Analysis (IPA) Version 01-13 (Qiagen, Germantown, MD) was used to discover pathways and
266 functions enriched among the DMR-associated genes based on Fisher Exact p-values of ≤ 0.05,
267 using the curated genes in IPA's Knowledgebase as the reference gene set. The overall workflow
268 for this study is summarized in **Figure 1**.
269

Class prediction analysis of WGBS data using machine-learning approaches

In order to select reliable predictors of DDE exposure levels for further validation by pyrosequencing, random forest (RF) and linear support vector machine (SVM) models were used to create binary classifiers that identify the exposure class (either 'First' or 'Third' tertile) of a given sample. A total of 261 DMRs from the analysis of all 52 samples, without ComBat batch correction, were utilized as predictors for the models. The samples were split according to sequencing batch, where the 32 samples from the first batch were used as a training set, and the 20 samples from the second batch were used as a testing set. The training set consisted of 16 samples in the 'Third' tertile and 16 matched controls labeled as the 'First' tertile. Similarly, the testing set consisted of 10 samples in each of these two tertile groups. Then, the machine learning model, either RF or SVM, was built with 5-fold cross-validation on the training set and used to predict the DDE exposure 'Tertile' class in the testing set.

To identify the most predictive DMRs for DDE exposure, we used two wrapper feature selection algorithms from the sigFeature Bioconductor package and the Boruta R package. Among the three types of feature selection methods (filter, embedded, and wrapper), we chose the wrapper method as it usually provides the most relevant feature set for a specific model. Wrapper methods rank features by repeatedly generating a subset of features from the full set of features and training a particular type of model using the generated subset.

There are many variable selection methods for the RF algorithm, but the Boruta algorithm was shown to be the most relevant, as it identifies all relevant variables, including variables that contain redundant information. In genomics studies, the inclusion of correlated and thus redundant variables may be important for model performance and the consistency of

292 results(56). The Boruta algorithm in the Boruta package applies the RF algorithm from the
293 randomForest R package (57).
294 Of the variable selection methods that use the SVM algorithm, we decided to use the
295 SVM recursive feature elimination (SVM-RFE) algorithm and t-statistic from the sigFeature
296 package. SVM-RFE, one of the most effective feature selection methods, uses a greedy
297 algorithm to find the best subset of features for binary classification tasks but does not consider
298 differentially significant features between classes. The sigFeature package addresses this
299 limitation by using the t-statistic to also find differentially significant features (58).
300 ***Pyrosequencing analyses***
301 Pyrosequencing analyses of selected DMRs were performed in the Weksberg laboratory
302 at The SickKids Research Institute of The Hospital for Sick Children; Toronto, Ontario, Canada.
303 These DMRs included specific regions for *CSMD1, NRXN2, RBFOX1, MIRLET7BHG,*
304 *PTPRN2, SNORD115-30,* and *SNORD115-37* that were identified by the WGBS analyses.
305 Primers for each of the regions that typically included multiple methylation sites as well as the
306 PCR conditions for the pyrosequencing analyses are provided in **Additional file 7**. The resulting
307 pyrosequencing data were returned to us for further analyses of the methylation profiles of each
308 region as a function of DDE serum concentration (µg/gm lipid) for each sample.
309 ***Hypergeometric distribution analyses***
310 Hypergeometric distribution analyses were employed to identify the significance of
311 overlap between DMR-associated genes and autism risk genes from the SFARI Gene database
312 (59). First, the overlapping genes were identified using a Venn diagram software program called
313 Venny 2.1.0 https://bioinfogp.cng.csic.es/tools/venny/ (60). Significant overlap between the
314 DMR-associated genes and SFARI genes was determined by hypergeometric distribution

analyses using the CASIO Keisan Online Calculator <http://keisan.casio.com/exec/system/1180573201>, with significance determined by an upper cumulative Q-value of ≤ 0.05. These two programs were also used to identify significant overlap of DMR-associated genes from this study and those from other studies.

Results

DMRs associated with high and low exposures to DDE

WGBS analyses of the initial 32 samples (discovery set) with correction for all covariates revealed a total of 894 differentially methylated regions (DMRs, permutation $p \leq 0.05$) between the first and third DDE exposure tertiles, while subsequent analyses of the validation set of 20 samples resulted in a total of 865 DMRs (**Figure 2,** DMRs in **Additional files 5 and 6**). The overall distribution profiles of annotated gene and CpG regions were similar for both analyses (**Additional file 8**). Analysis of larger blocks of sequence for differential methylation revealed a single region that reached genome-wide significance after multiple testing correction ($q < 0.022$) when all 52 samples were combined (**Figure 3**). This block covered a 40,618 bp region of the *SNORD115* locus that is maternally imprinted, meaning expressed exclusively from the paternal allele. The validation data set shows nominal significance for increased methylation across this locus with a width of 63,606 bp ($p < 0.011$).

Sperm DMRs associated with neurodevelopmental processes and ASD-risk genes

Gene ontology analyses of the genes within the DMRs in both discovery and validation samples were performed using two different gene mapping and ontology approaches. GOfuncR analysis provides an overview of the top biological processes, cellular components, and molecular functions associated with the DMR-associated genes from both discovery and validation sets (**Figure 4**), while STRING analysis of these genes not only replicates some of the

GO terms from the GOfuncR analysis but also reveals significant over-representation of a number of processes involved in nervous system development and function that are shared by both datasets (**Table 1**). Notably, these shared processes include nervous system development, generation of neurons, neurogenesis, neuron differentiation, and synapse organization.

Table 1. Gene ontology terms enriched among DDE DMR-associated genes from discovery and validation sets

GO ID	Pathway description	Discovery FDR	Validation FDR
GO.0007399	nervous system development	1.80E-11	4.80E-03
GO.0007275	multicellular organism development	2.13E-08	1.60E-03
GO.0048731	system development	3.14E-08	4.10E-03
GO.0048856	anatomical structure development	7.91E-08	2.20E-03
GO.0022008	neurogenesis	5.51E-07	2.19E-02
GO.0032502	developmental process	5.51E-07	5.50E-04
GO.0048699	generation of neurons	5.51E-07	1.17E-02
GO.0048666	neuron development	1.55E-06	4.12E-02
GO.0030182	neuron differentiation	2.95E-06	2.19E-02
GO.0000904	cell morphogenesis involved in differentiation	1.03E-05	3.89E-02
GO.0048468	cell development	2.77E-05	1.48E-02
GO.0030154	cell differentiation	1.50E-04	2.19E-02
GO.0048869	cellular developmental process	2.50E-04	1.45E-02
GO.0048513	animal organ development	2.80E-04	1.92E-02
GO.0003279	cardiac septum development	7.10E-04	3.89E-02
GO.0050808	synapse organization	1.80E-03	4.18E-02
GO.0072359	circulatory system development	1.80E-03	4.12E-02
GO.0007155	cell adhesion	2.40E-03	9.60E-03
GO.0003148	outflow tract septum morphogenesis	3.30E-03	2.74E-02
GO.0050793	regulation of developmental process	4.10E-03	2.60E-02
GO.0050794	regulation of cellular process	5.80E-03	1.90E-04
GO.0000122	negative regulation of transcription by RNA polymerase II	6.10E-03	7.70E-04
GO.2000026	regulation of multicellular organismal development	6.10E-03	3.89E-02
GO.0050789	regulation of biological process	1.04E-02	3.30E-04
GO.0006928	movement of cell or subcellular component	1.50E-02	2.19E-02
GO.0007423	sensory organ development	1.50E-02	1.05E-02
GO.0048518	positive regulation of biological process	3.04E-02	3.80E-04
GO.0050767	regulation of neurogenesis	4.36E-02	4.12E-02
GO.0051960	regulation of nervous system development	4.54E-02	3.51E-02

The enrichment in neuronal processes was further confirmed by pathway and functional analyses of the DMR-associated genes using IPA. CREB signaling in neurons, the endocannabinoid

352 **Table 2. Canonical pathways enriched among DDE DMR-associated genes from discovery**
353 **and validation sets**
354

Canonical Pathways (Discovery)	-log(p-value)*	Molecules
CREB Signaling in Neurons	6.18	CACNA1I,CACNG6,GRID2,GRIA1,GRIK3,GNG2,FLT3,CACNB4,PIK3C2G,GNAI1,FGFR2,CREB5,GNG7,CACNA1A,GRM5,SHC1,GNAO1,IRS2,PRKCH,GRIK2,GRIK1
Glutamate Receptor Signaling	3.85	GRM5,GRID2,GRIA1,GNG2,GRIK3,GRIK2,GNG7,GRIK1
Endocannabinoid Developing Neuron Pathway	3.65	FLT3,GNAO1,GNG2,GNAI1,PIK3C2G,FGFR2,PAX6,IRS2,GSK3B,CREB5,CTNNB1,GNG7
Netrin Signaling	3.4	CACNA1I,NCK2,CACNG6,UNC5A,CACNB4,RYR3,NFATC1,CACNA1A
Calcium Signaling	2.91	CACNA1I,CACNG6,MYH10,HDAC4,MYH13,GRIA1,CACNB4,TRPC4,CREB5,NFATC1,CACNA1A,RYR3,SLC8A1,GRIK1
G Beta Gamma Signaling	2.77	CACNA1I,CACNG6,SHC1,GNAO1,CACNB4,GNG2,GNAI1,PRKCH,GNG7,CACNA1A
Integrin Signaling	2.6	TSPAN5,ITGA8,FLT3,PIK3C2G,TSPAN2,FGFR2,TNK2,ITGAL,NCK2,SHC1,IRS2,GSK3B,CTTN,NEDD9
Axonal Guidance Signaling	2.42	MMP20,LRRC4C,UNC5A,NTN4,GNG2,FLT3,PIK3C2G,GNAI1,SEMA6B,FGFR2,SLIT2,DPYSL5,PDGFC,GNG7,NFATC1,NCK2,SHC1,SEMA6D,NTNG2,GNAO1,IRS2,PRKCH,GSK3B,SEMA3C
Androgen Signaling	2.37	CACNA1I,CACNG6,SHC1,GNAO1,CACNB4,GNG2,GNAI1,PRKCH,GNG7,CACNA1A
Relaxin Signaling	2.31	PDE10A,RXFP1,FLT3,GNAO1,GNG2,GNAI1,PIK3C2G,FGFR2,IRS2,PDE4B,GNG7
Growth Hormone Signaling	2.02	FLT3,CSHL1,PIK3C2G,FGFR2,CSH1/CSH2,IRS2,PRKCH
Gap Junction Signaling	1.95	GJA10,GRIA1,FLT3,GRIK3,GNAI1,PIK3C2G,FGFR2,IRS2,PRKCH,GRIK2,CTNNB1,GRIK1
Synaptic Long Term Depression	1.94	GRM5,CACNA1I,CACNG6,GRIA1,GRID2,RYR3,GNAO1,CACNB4,GNAI1,PRKCH,CACNA1A
G-Protein Coupled Receptor Signaling	1.94	HTR5A,PDE10A,FLT3,PIK3C2G,GNAI1,FGFR2,DRD5,PDE4B,CREB5,CHRM3,GRM5,SHC1,GNAO1,IRS2,DUSP4
GABA Receptor Signaling	1.82	CACNA1I,CACNG6,KCNN3,GABRB3,CACNB4,SLC6A1,CACNA1A
Huntington's Disease Signaling	1.68	GRM5,SHC1,HDAC4,IFT57,FLT3,GNG2,PIK3C2G,FGFR2,DNM3,IRS2,PRKCH,CREB5,GNG7
Neurotrophin/TRK Signaling	1.55	SHC1,FLT3,PIK3C2G,FGFR2,IRS2,CREB5
α-Adrenergic Signaling	1.42	GNG2,GNAI1,PRKCH,PYGL,SLC8A1,GNG7
Canonical Pathways (Validation)	**-log(p-value)***	**Molecules**
CREB Signaling in Neurons	4.15	RAP2B,CACNG6,CACNA1H,PIK3R5,GNAI1,GNG7,GRM5,ADCY9,CACNA2D1,ADCY1,PRKAR1B,PIK3R6,ATF4,IRS2,CACNB2,ADCY8,GNAL,GRIA3
Endocannabinoid Developing Neuron Pathway	4.02	RAP2B,PIK3R5,GNAI1,GNG7,ADCY9,ADCY1,PRKAR1B,PIK3R6,ATF4,IRS2,ADCY8,CTNNB1,GNAL
Calcium Signaling	3.14	RAP2B,CACNG6,HDAC4,MYH9,MYH14,HDAC1,CACNA1H,TRPC7,CACNA2D1,MYH3,PRKAR1B,ATF4,CACNB2,CHRNA3,GRIA3
GABA Receptor Signaling	2.77	GABRG3,CACNG6,ADCY9,ADCY1,CACNA2D1,GABRA6,CACNA1H,CACNB2,ADCY8
G Beta Gamma Signaling	2.6	RAP2B,CACNG6,ADCY1,CACNA2D1,GNAI1,PRKAR1B,CACNA1H,CACNB2,GNG7,GNAL
Gap Junction Signaling	2.56	RAP2B,GJA1,ACTB,PIK3R5,GNAI1,ADCY9,ADRB1,ADCY1,PRKAR1B,PIK3R6,IRS2,CTNNB1,ADCY8,GRIA3
Synaptic Long Term Depression	2.18	RAP2B,GRM5,CACNG6,CACNA2D1,GNAI1,CACNA1H,CACNB2,PPP2R5C,PPP2R5E,NOS2,GNAL,GRIA3
Netrin Signaling	1.94	CACNG6,CACNA2D1,PRKAR1B,CACNA1H,CACNB2,UNC5C
Notch Signaling	1.66	MAML2,HES7,JAG1,PSEN1
Ephrin Receptor Signaling	1.48	ITGB1,RAP2B,EPHA6,SDCBP,SH2D3C,GNAI1,ATF4,EPHA3,GNG7,GNAL
Serotonin Receptor Signaling	1.48	ADCY9,SMOX,ADCY1,ADCY8
Axonal Guidance Signaling	1.37	RAP2B,ITGB1,BMP4,NRP2,ADAMTS20,PTCH1,PIK3R5,GNAI1,EPHA3,ROBO1,GNG7,EPHA6,SDCBP,WNT3A,PIK3R6,PRKAR1B,IRS2,BMP6,MMP17,GNAL,UNC5C

*Negative logarithm of the Fisher exact p-value indicating the probability that the indicated function is not enriched among the indicated genes based on the reference set of genes in the IPA Knowledgebase

355
356

developing neuron pathway, netrin signaling, calcium signaling, GABA signaling, and axon guidance signaling are among the canonical pathways significantly enriched in both data sets (**Table 2**), while recognition of neurons, outgrowth of neurites, and neurotransmission are shared functions over-represented among the DMR-associated genes (**Table 3**).

Table 3. Nervous system functions enriched among DDE DMR-associated genes from discovery and validation sets

Nervous system development and functions (Discovery)	**p-value***	**Molecules**
Development of central nervous system	3.71E-06	ANKLE2,ASIC2,ATOH1,CNTN6,CNTNAP2,EML1,GRIK1,GSK3B,HGF,JARID2,MBP,MYO16,PAX6,PDGFC,PROX1,TRAPPC9
Synaptic transmission	1.22E-03	ASIC2,GRIA1,GRIK1,GRIK2,GRM5,MBP,NRG3,RIT2,SLC6A1,SYT1
Recognition of neurons	2.23E-03	NTM,OPCML
Outgrowth of neurites	3.11E-03	GFRA2,GSK3B,HGF,mir-124,SHC1,SLIT2,TGFA
Guidance of axons	3.88E-03	DOK5,GFRA2,IRS2,NRXN3,NTN4,SHC1,SLIT2,UNC5A
Generation of nervous tissue cell lines	7.17E-03	MYT1L,RMST
Outgrowth of axons	1.06E-02	HGF,SLIT2
Formation of brain	1.83E-02	CNTNAP2,EML1,GSK3B,HGF,MYO16,TRAPPC9
Nervous system development and functions (Validation)	**p-value***	**Molecules**
Recognition of neurons	2.50E-03	NTM,OPCML
Outgrowth of neurites	4.23E-03	BMP4,ITGA1,ITGB1,mir-10,TGFA,TIAM1,WNT3A
Proliferation of neuronal cells	5.08E-03	BMP4,ITGA1,ITGB1,JAG1,mir-10,TGFA,TIAM1,WNT3A
Neurotransmission	9.48E-03	CBLN1,DTNA,GPR176,GRM5,KCNQ1,MBP,MYH14,PSEN1,SYT1
Quantity of neuroepithelial cells	1.18E-02	BMP4,BMP6

*Fisher exact p-value indicating the probability that the indicated function is not enriched among the indicated genes based on the reference set of genes in the IPA Knowledgebase

To further investigate the relevance of these DMR-associated genes to ASD, we performed hypergeometric distribution analyses to determine enrichment in autism risk genes from the SFARI Gene database. **Figure 5** shows the overlap between DMR-associated genes from the discovery and validation analyses as well as the overlap between these DMR associated genes and the SFARI genes. The upper cumulative Q-values for enrichment in SFARI genes were 2.5×10^{-7} and 1.1×10^{-5} for the discovery and validation sets, respectively. Of the 138

overlapping DMR-associated genes between the discovery and validation sets, 14 are included in the SFARI Gene database. Notably, all of these genes are involved in development.

Class prediction analysis using a machine-learning approach

An attempt was made to identify DMRs that could reliably assign samples to the high or low exposure levels. Both random forest and linear support vector machine models were independently applied to the entire set of 261 DMRs obtained using all 52 samples (**Additional file 9**). The discovery set of samples was used for training the classifier, and the validation set was used for testing the classifier. **Table 4** shows 15 of the top predictors resulting from these two classification approaches.

Table 4. Top classifier genes that predict exposure level from a combination of two separate machine-learning analyses

Chr	start	end	annotation	distance to TSS	Gene Symbol	Gene Name
chr2	98280512	98281225	Intron 23 of 28	193396	VWA3B	von Willebrand factor A domain containing 3B
chr2	240961395	240961697	Intron 20 of 31	5754	LOC200772	uncharacterized LOC200772
chr3	17133705	17133973	Distal Intergenic	123211	PLCL2	phospholipase C like 2
chr3	43484849	43485664	Intron 11 of 12	136404	ANO10	anoctamin 10
chr4	2161268	2161626	Intron 11 of 23	67605	POLN	DNA polymerase nu
chr5	7927007	7928271	Distal Intergenic	30570	MTRR	5-methyltetrahydrofolate-homocysteine methyltransferase reductase
chr9	65389175	65390088	Distal Intergenic	-103966	FOXD4L5	forkhead box D4 like 5
chr10	14980617	14981682	Intron 1 of 2	8062		??
chr10	128066899	128068111	Intron 11 of 20	159838	PTPRE	protein tyrosine phosphatase, receptor type E
chr10	14989759	14992906	Distal Intergenic	17204		??
chr13	110824201	110825915	Distal Intergenic	-11117	LINC00567	long intergenic non-protein coding RNA 567
chr13	114042350	114043198	Intron 3 of 23	89413	RASA3	RAS p21 protein activator 3
chr20	63948311	63949065	Exon 2 of 3	7350	UCKL1	uridine-cytidine kinase 1 like 1
chr22	46103014	46103671	Exon 4 of 5	49145	MIRLET7BHG	MIRLET7B host gene
chr22	46104978	46105643	Exon 5 of 5	51109	MIRLET7BHG	MIRLET7B host gene

388 ***Pyrosequencing analyses of several DMR-associated ASD risk genes***

389 Several DMRs were selected for pyrosequencing validation. These included the regions
390 harboring known autism risk genes, *CSMD1*, *NRXN2*, and *RBFOX1*, all of which were found to
391 be differentially methylated in both discovery and validation analyses. In addition, we included
392 *PTPRN2* which, like *NRXN2*, was found to be differentially methylated in cord blood from the
393 Faroese population as a function of level of DDE exposure (61). Also selected for
394 pyrosequencing were *MIRLET7HBG* which was implicated by the classifier analyses and two
395 SNORDs (*SNORD115-30* and *SNORD115-37*) which are located in an imprinted region
396 identified as differentially methylated in this study as well as in a previous study on paternal
397 sperm from men with an autistic child (43). **Figure 6** shows the correlation curves for
398 methyation level versus serum DDE exposure level for *CSMD1*, *NRXN2*, and *RBFOX1*. All of
399 these genes showed a modest but statistically significant inverse correlation between methylation
400 and DDE exposure levels, whereas *MIRLET7HBG* showed a trend towards increased
401 methylation with increasing DDE levels (**Figure 7**). Similarly, consistent but not significant
402 increases in methylation at multiple CpG sites within *SNORD115* were observed in the samples
403 with higher exposures (**Table 5**). *PTPRN2*, on the other hand, showed no correlation between
404 methylation detected by pyrosequencing and exposure levels.

405 ***Comparison of DMR-associated genes from this study with those from ASD-related and***
406 ***unrelated methylation studies***

407 The DDE DMR-associated genes in sperm were compared to differentially methylated
408 genes in a variety of tissues from studies on ASD-associated methylation differences. These
409 tissues included cord blood from newborns, a fraction of whom was later diagnosed with ASD
410 (62), sperm (43) and placenta (63) from parents of children with high risk for ASD, brain tissues

from individuals with Dup15q syndrome that is often associated with ASD (27) as well as lymphoblastoid cell lines derived from individuals with a severe form of ASD (64).

Table 5. Differential methylation of *SNORD115-30* and *SNORD115-37* by DDE tertile validated by pyrosequencing analyses

SNORD115-30*								Average
Tertile	Pos. 1	Pos. 2	Pos. 3	Pos. 4	Pos. 5	Pos. 6	Pos. 7	All Positions
First	28.98	30.76	34.31	33.49	37.26	44.31	30.56	34.24
Third	33.82	35.65	39.14	37.78	39.80	47.28	33.46	38.13
Difference (Third-First)	4.84	4.89	4.82	4.30	2.54	2.97	2.89	3.89

*Methylation was quantified at 7 specific positions in the SNORD115-30 DMR

SNORD115-37♦					Average
Tertile	Pos. 1	Pos. 2	Pos. 3	Pos. 4	All positions
First	41.12	36.29	33.84	41.63	38.22
Third	45.86	41.88	39.41	46.01	43.29
Difference (Third-First)	4.74	5.59	5.57	4.38	5.07

♦Methylation was quantified at 4 specific positions in the SNORD115-37 DMR

Hypergeometric distribution analyses show that the DMR-associated genes from both the discovery and validation WGBS analyses overlapped significantly with those from ASD cord blood, paternal sperm, placenta, and Dup15q brain that also showed detectable PCB 95 exposures (27) (**Table 6**). DMR-associated genes from the discovery, but not validation, set also overlapped significantly with those from lymphoblastoid cell lines. Interestingly, the significance of the overlap between the DDE DMR-associated genes is highest in DMRs identified from cord blood from children diagnosed with ASD and lower in the more differentiated or immortalized tissues, i.e., brain and lymphoblasts, respectively. In the study on placental methylation, placentas of high-risk mothers with a child (or children) already diagnosed with ASD were obtained after the birth of a subsequent child for WGBS analysis (65). Subsequent analysis of DMRs from the placentas for ASD outcome revealed 596 nominally significant genes, of which two (*CYP2E1* and *IRS2*) reached genome-wide significance (63). For the sperm methylation

Table 6. Overlap among DDE DMR-associated genes and DMRs from different ASD studies and tissues

Samples for comparison of DDE DMRs (sperm)		Hypergeometric distribution Q-value	
(# DMR-associated genes)		Discovery (742)	Validation (763)
Cord blood from newborns later diagnosed with ASD[62]	(2173)	2.96E-23	8.56E-10
Sperm (high ASD risk)[43]	(144)	1.22E-06	9.95E-05
Placenta (ASD outcome)[63]	(596)	1.56E-04	6.78E-05
Brain (Dup15q)[27]	(942)	1.90E-03	2.00E-03
Lymphoblastoid cell lines[64,*]	(827)	1.90E-02	8.00E-02
Pan-Cancer[66,♦]	(434)	7.34E-12	2.17E-09
Faroese cord blood (DDE associated)[61]	(1136)	0.608	0.55
[62] Mordaunt et al., 2020			
[43] Feinberg et al., 2015			
[63] Zhu et al., 2019			
[27] Dunaway et al., 2016			
[64] Hu et al., 2020; *males only, severely language-impaired			
[66] Su et al., 2018; ♦ hypermethylated canyon genes			
[61] Leung et al., 2018			

analysis, semen samples from fathers of a child already diagnosed with ASD, were obtained for methylation analyses (43). The placenta and sperm studies revealed DMR-associated genes related to ASD outcome or higher risk for ASD, respectively, in offspring of the high-risk parents in comparison to parents of neurotypical children, while the DMR-associated genes in cord blood may be more directly related to ASD diagnosis in the individual. In addition, the DDE DMR-associated genes from both discovery and validation datasets significantly overlapped with those from pan-cancer studies (66). The overlap with cancer genes is not surprising since organochlorine exposures are well-known risk factors for cancer; thus, the current study reveals exposure-associated epigenetic changes in DMR-associated genes that may also increase cancer risk (67-69). Interestingly, cancer and ASD share many risk genes and pathways (70) , some of which may be influenced by environmental factors. On the other hand,

there was no significant overlap of DDE DMR-related genes with those in cord blood from the Faroese population that were also associated with DDE exposures (61), although DMR-associated genes from both sperm (discovery set) and cord blood showed enrichment in multiple canonical pathways associated with neurological functions (**Table 7**). Comparison of the DMR-

Table 7. Canonical pathways enriched among overlapping DDE DMR-associated genes between Faroese sperm and cord blood

Canonical Pathways	-log(p-value)*	Molecules
Netrin Signaling	3.53	CACNA1I,NCK2,UNC5A
GABA Receptor Signaling	3.06	CACNA1I,GABRB3,KCNN3
Circadian Rhythm Signaling	2.71	CREB5,VIP
CREB Signaling in Neurons	2.02	CACNA1I,CREB5,GRIK2
Axonal Guidance Signaling	1.82	NCK2,NTN4,NTNG2,UNC5A
Cellular Effects of Sildenafil (Viagra)	1.58	KCNN3,PDE4B
GP6 Signaling Pathway	1.55	COL25A1,FGA
Corticotropin Releasing Hormone Signaling	1.51	CACNA1I,CREB5
Extrinsic Prothrombin Activation Pathway	1.5	FGA
Gustation Pathway	1.44	CACNA1I,PDE4B
Methionine Degradation I (to Homocysteine)	1.41	MGMT
Protein Kinase A Signaling	1.39	CREB5,PDE4B,PTPRS
Cysteine Biosynthesis III (mammalia)	1.37	MGMT
GNRH Signaling	1.35	CACNA1I,CREB5
Ephrin Receptor Signaling	1.31	CREB5,NCK2

*Negative logarithm of the Fisher exact p-value indicating the probability that the indicated pathway is not enriched among the indicated genes based on the reference set of genes in the IPA Knowledgebase

associated genes from the ASD and Faroese cord blood studies showed no significant overlap (Q = 0.84), but both sets of DMRs were highly enriched in genes on the X-chromosome. Intriguingly, the X-linked genes from the ASD cord blood were predominantly found in females (62), while those from the Faroese cord blood were exclusively male specific (61). By comparison, despite the highly significant overlap between the DMR-associated genes in the Faroese sperm and the ASD cord blood, there were relatively few X-linked DMR-associated genes found in sperm. Collectively, these results suggest that exposure to persistent organic

462 pollutants, such as DDE, is associated with altered methylation status of genes in sperm and
463 early developmental tissues that are critically associated with ASD.
464
465 **Discussion**
466
467 ***Methylation patterns of sperm DNA are associated with DDE exposure levels***
468 This study shows that the DNA methylation status in sperm may be influenced by
469 lifelong exposure to environmentally derived persistent EDCs, such as DDE. The Faroese cohort
470 used in this study is particularly exposed to higher than average levels of EDCs as a result of
471 their natural diet which includes substantial amounts of pilot whale meat and blubber. Fatty
472 tissues are reservoirs for lipophilic molecules, which include a wide variety of organochlorines,
473 such as DDE as well as PCBs. The high correlation between the levels of multiple
474 organochlorines and DDE in serum indicates that the DDE exposures employed in this study are
475 proxies for exposures to persistent organochlorines in general. Although some of these
476 compounds have now been banned for use, the long half-lives of such compounds and/or their
477 breakdown products still pose risk of environmental exposures.
478 ***DMRs harbor genes enriched for neurodevelopment and function***
479 Although there are hundreds of genes whose methylation is altered by elevated exposure
480 to DDE and other organochlorines, our gene ontology and pathway analyses reveal that genes
481 involved in nervous system development and function are among the most significantly over-
482 represented in DMRs. Moreover, a significant number of these genes are also autism risk genes
483 that are included in the SFARI Gene database. Among the ASD-risk DMR-associated genes
484 validated by pyrosequencing are *CSMD1*, *NRXN2*, and *RBFOX1*. *CSMD1*, which encodes for
485 CUB and Sushi multiple domains 1, is highly expressed in brain tissues where it has been
486 associated with neuronal growth cone stabilization and neuritogenesis (71). Aside from being a

risk gene for ASD (72-74), it has also been implicated in schizophrenia, bipolar disorder, and post-traumatic stress disorder (75). *NRXN2* codes for neurexin 2, a brain-enriched cell adhesion molecule that has long been associated with ASD (76-78). NRXN2 plays a role in early cortical synaptogenesis and axon guidance (79). *RBFOX1* encodes for RNA binding fox-1 homolog 1, a neuron-specific splicing factor that has been implicated in many studies on ASD (80-82). Interestingly, *RBFOX1* was one of the top transcriptional targets of the orphan nuclear hormone receptor RORA which we found to be regulated by sex hormones (12) as well as EDCs, including DDE and atrazine, an herbicide (Shu, Kocher, and Hu, unpublished data).

EDC-associated differential methylation also impacts noncoding regions of the genome

MIRLET7BHG codes for a long noncoding RNA (lncRNA) whose neonatal umbilical cord methylation level has been associated with birth weight in a study of the effects of prenatal environment and genotype on offspring weight and obesity in early childhood (83). This gene was identified in this study as a potential predictor of exposure level by machine-learning analyses. Interestingly, this gene was one of eight lncRNAs that were differentially expressed between a group of women with polycystic ovary syndrome (PCOS) and control women, and the only one whose expression was correlated with BMI (84). As PCOS is associated with abnormally high levels of male hormones, it is notable that the methylation status of *MIRLET7BHG* is influenced by EDC exposures.

The altered methylation in a large block on chromosome 15 encompassing the noncoding *SNORD115* region is of particular interest inasmuch as this region was also found to be differentially methylated in the sperm of fathers with a child exhibiting ASD in comparison to the sperm of fathers of unaffected children (43). However, the origins of such differences in sperm DNA methylation that are associated with ASD are unknown. The results from this study

suggest that environmental exposures to certain EDCs may in part be responsible for alterations in the sperm methylome and, in particular, the *SNORD115* region. This region is imprinted maternally and normally shows compact heterochromatin until its expression, which is exclusive to neurons. Intriguingly, the *SNORD115* locus is not expressed in any tissue besides the brain. With respect to function, the *SNORD115* region on human chromosome 15q11-q13 contains a series of 48 highly conserved small nucleolar RNAs (snoRNAs) that participate in the modification of other noncoding RNAs and site-specific 2′-O-methylation of substrate RNAs (85) as well as alternative splicing and RNA editing, especially of 5-HTR2C pre-mRNA (86-88). In addition, a previous study has also identified alterations at this locus in association with exposures to EDCs, albeit of a more transient (non-persistent) nature. Specifically, increased methylation was observed in the *SNORD115* locus in human fetal lung tissue of discontinued pregnancies of women exposed to BPA (89).

Although deletions in this genomic region are associated with Prader-Willi syndrome (PWS), loss of *SNORD115* alone is not sufficient to cause the disease (90,91). Maternal duplications in the chr15q11-q13 region (aka. Dup15q syndrome) have also been associated with ASD as well as other developmental disorders, with some genes showing altered methylation status (92-94). As mentioned earlier, Dup15q was shown to be a strong predictor of PCB 95 exposure (28). The Mitchell et al. study further showed that LINE-1 methylation was reduced in Dup 15q and PWS samples but not idiopathic ASD, suggesting gene x environment interactions possibly mediated through epigenetic modifications in the genetically defined but not idiopathic ASD. However, *SNORD115* genes were not specifically implicated in these studies. Thus, the present study, coupled with that of Feinberg et al. on sperm from fathers of a child with ASD (43), reveals an additional epigenetic mechanism through which long-lived EDCs may mediate

ASD-related changes in this region. Moreover, such changes in germline cells raise the possibility of transgenerational inheritance of phenotype as described in multiple animal studies.

Potential significance of methylation changes in sperm cells

Previous studies on animal models have reported transgenerational effects of EDC exposures on disease and behavioral phenotypes, some of which were shown to be mediated by epigenetic changes in the germline (25,26,95,96). Initial findings in rodents have shown that exposure to vinclozolin (an androgenic EDC) or methoxychlor (an estrogenic EDC) led to increased male infertility and related characteristics such as decreased sperm count in all subsequent generations, from F1 through F4 (97). In addition, when gestating F0 females were given various doses of a mix of EDCs during embryonic development, F1 and F3 generations exhibited increased total disease. Differentially methylated regions were found in sperm of the the F3 generation that included promoters of genes associated with underlying diseases such as obesity, PCOS, and ovarian disease (98). These studies and others (26) indicate that F0 exposure can lead to transgenerational effects on phenotypes that were associated with epigenetic changes in sperm that were specific to adult onset disease.

Advantages and limitations of this study design and future directions

Although an obvious limitation of this study is the relatively low number of samples, this limitation is somewhat offset by the fact that the Faroe Islands population is considered a "genetic isolate", thus reducing genetic heterogeneity that is often a major challenge in epigenetic studies. A further advantage is the natural Faroese diet, rich in pilot whale meat and blubber, that exposes the population to higher than average levels of persistent organochlorines. Another limitation of this study is the lack of a completely unexposed set of samples for comparisons since low levels of persistent EDCs can be detected in a majority of individuals.

Because of the non-monotonic dose-response behavior of EDCs (99), the lowest exposures examined here may show even greater methylation differences relative to truly unexposed controls than the differences between the first and third exposure tertiles included in this study. Moreover, we were not able to correlate methylation changes with changes in gene expression in the same tissues inasmuch as spermatocytes are transcriptionally inactive. Additionally, we could not correlate EDC exposure with ASD risk in offspring, as there was no information on children (if any) of the young men in this Faroese cohort. Such a study would be particularly interesting as neurobehavioral deficits in a Faroese birth cohort of 7-year-old children have been associated with prenatal exposures to organochlorine neurotoxicants in seafood as measured in umbilical cord tissue (45). In addition to revealing relationships between organochlorine exposures, sperm DNA methylation profiles, and neurodevelopmental disorders, the availability of this publicly accessible methylation data on DDE/organochlorine-associated changes in the sperm methylome will provide a valuable resource for studies on other diseases and conditions, such as cancer, obesity, diabetes, and infertility, which are also linked to environmental exposures.

Future studies should include additional and larger cohorts sampled longitudinally to monitor temporal changes in individual sperm methylation levels as a function of cumulative EDC exposures as well as expanded concentration levels of DDE. Given the association between exposure levels of DDE (which were highly correlated with the sum of the most prevalent long-lived PCBs) and altered methylation of many neurodevelopmental genes, it will also be of interest to investigate the relationships between DDE/organochlorine exposures of the men, epigenetic changes in sperm, and the health outcomes of their children.

Conclusions

This study shows that elevated exposure to DDE, one of a class of persistent organochlorines, is associated with differential genome-wide DNA methylation patterns in sperm when compared against the lowest measurable exposure levels. The DMRs are enriched for genes involved in many biological processes, including neurological functions and pathways impacted by neurodevelopmental disorders. This study thus supports the link between environmental EDC exposures and epigenetic changes in germ cells that may impact the regulation of many genes associated with disease phenotypes, including ASD. Studies involving animal models have shown that DNA methylation patterns as well as associated phenotypes or diseases can be stably and heritably transmitted through the germline in several generations of offspring, specifically in relation to EDC exposure. It is unknown whether the DNA methylation differences noted in this study arose from direct exposure of the individual sperm donors to EDCs or from ancestral exposures that led to the inheritance of specific DNA methylation patterns in the donors' sperm.

List of abbreviations

ASD: autism spectrum disorder

BMI: body mass index

BPA: bisphenol A

DDE: 1,1-dichloro-2,2-bis(p-chlorophenyl)ethylene

DDT: dichloro-diphenyl-trichloroethane

DMR: differentially methylated region

EDC: endocrine disrupting compound

GEO: Gene Expression Omnibus

GO: gene ontology

IPA: Ingenuity Pathway Analysis

LCL: lymphoblastoid cell lines

PBDE: polybrominated diphenyl ether

PCB: polychlorinated biphenyl

PCOS: polycystic ovary syndrome

PCR: polymerase chain reaction

POP: persistent organic pollutant

PWS: Prader-Willi syndrome

RF: random forest

SFARI: Simons Foundation Autism Research Initiative

SVM: support vector machine

SVM-RFE: SVM-recursive feature elimination

WGBS: whole genome bisulfite sequencing

Declarations

Ethics approval and consent to participate: This study, which uses de-identified semen samples, was not considered as "human subject research" by GWU's IRB committee for Human Subject Research. The initial collection of semen was performed by the Faroese Hospital System under the direction of Dr. Pál Weihe for a parent study originally approved by the local Science Ethics Committee for the Faroe Islands.

Consent for publication: Not applicable

Availability of data and materials: All raw and processed data from the WGBS analyses have been deposited into GEO (GEO Accession #: GSE165915).

Competing interests: All authors declare no competing interests.

Funding: This study was supported by the National Institute of Environmental Health Sciences of the National Institutes of Health (grant R21 ES028124 to VWH and partial support from R01 ES029213 to JML). BIL was supported by a Canadian Institutes of Health Research (CIHR) postdoctoral fellowship (MFE-146824) and a CIHR Banting postdoctoral fellowship (BPF-162684). None of the funding agencies played any role in the design of the study, collection, analysis, and interpretation of the data, and in the writing of the manuscript.

Authors' contributions

AGM was responsible for sequence alignment, quality control, WGBS, and bioinformatics analyses of the DMRs. HTS was responsible for sample organization and preparation, including sperm and DNA isolation from the semen aliquots. BIL also contributed substantially to WGBS and bioinformatics analyses. HH performed the machine learning analysis. CB analyzed the pyrosequencing data. YL provided statistical support and advice. JML provided advice and discussion of the WGBS analysis and data interpretation. VWH conceived of the study, performed bioinformatics and hypergeometric analyses on the DMR-associated genes, and wrote the manuscript. AGM, BIL, and JML also contributed to manuscript preparation and editing. All authors read and approved the final manuscript.

Acknowledgements

We thank Dr. Pál Weihe (Department of Occupational Medicine and Public Health, Faroese Hospital System, Tórshavn, Faroe Islands) for generously providing the semen samples, demographic information on the semen donors, and data on serum levels of DDE and other organochlorines that had been determined previously by Dr. Philippe Grandjean (Department of Environmental Medicine, University of Southern Denmark, Odense, Denmark). We also thank Dr. Melissa Perry (Department of Environmental and Occupational Health, GWU) for dividing the semen samples from the Faroese cohort into exposure tertiles based on DDE serum levels which allowed us to process the samples upon receipt from sperm isolation through bisulfite sequencing without knowledge of the donor's exposure group.

References

(1) American Psychiatric Association. Task Force on DSM-5. Diagnostic and Statistical Manual of Mental Disorders. 5th ed. Arlington, VA: American Psychiatric Association Publishing; 2013.

(2) Maenner MJ, Shaw KA, Baio J, Washington A, Patrick M, DiRienzo M, et al. Prevalence of autism spectrum disorder among children aged 8 Years-Autism and developmental disabilities monitoring network, 11 Sites, United States, 2016. MMWR Surveill Summ 2020;69(4):1-12.

(3) Nevison CD. A comparison of temporal trends in United States autism prevalence to trends in suspected environmental factors. Environ Health Global Access Sci Sour 2014;13(1).

(4) Hallmayer J, Cleveland S, Torres A, Phillips J, Cohen B, Torigoe T, et al. Genetic heritability and shared environmental factors among twin pairs with autism. Arch Gen Psychiatry 2011;68(11):1095-1102.

(5) Tordjman S, Somogyi e, Coulon N, Kermarrec S, Cohen D, Bronsard G, et al. Gene X Environment Interactions in Autism Spectrum Disorders: Role of Epigenetic Mechanisms. Frontiers in Psychiatry 2014; 5:53.

(6) Grafodatskaya D, Chung B, Szatmari P, Weksberg R. Autism Spectrum Disorders and Epigenetics. J Am Acad Child Adolesc Psychiatry 2010; 49(8):794-809.

(7) Rangasamy S, D'Mello SR, Narayanan V. Epigenetics, Autism Spectrum, and Neurodevelopmental Disorders. Neurotherapeutics 2013;10(4):742-756.

(8) 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 Aug;24(8):3036-3051.

(9) Hu VW, Frank BC, Heine S, Lee NH, Quackenbush J. Gene expression profiling of lymphoblastoid cell lines from monozygotic twins discordant in severity of autism reveals differential regulation of neurologically relevant genes. BMC Genomics 2006;7:118.

(10) Hu VW, Nguyen A, Kim KS, Steinberg ME, Sarachana T, Scully MA, et al. Gene expression profiling of lymphoblasts from autistic and nonaffected sib pairs: altered pathways in neuronal development and steroid biosynthesis. PLoS One 2009;4(6):e5775.

(11) Sarachana T, Xu M, Wu R-C, Hu VW. Sex hormones in autism: Androgens and estrogens differentially and reciprocally regulate RORA, a novel candidate gene for autism. PLoS ONE 2011;6(2):e17116.

(12) Sarachana T, Hu VW. Genome-wide identification of transcriptional targets of RORA reveals direct regulation of multiple genes associated with autism spectrum disorder. Molecular Autism 2013;4(1):14.

(13) Baron-Cohen S, Knickmeyer RC, Belmonte MK. Sex Differences in the Brain: Implications for Explaining Autism. Science 2005;310(5749):819-23.

(14) Baron-Cohen S, Auyeung B, Nørgaard-Pedersen B, Hougaard DM, Abdallah MW, Melgaard L, et al. Elevated fetal steroidogenic activity in autism. Mol Psychiatry 2015;20(3):369-76.

(15) Hu VW. Is retinoic acid-related orphan receptor-alpha (RORA) a target for gene-environment interactions contributing to autism? Neurotoxicology 2012;33(6):1434-1435.

(16) Braun JM. Early-life exposure to EDCs: Role in childhood obesity and neurodevelopment. Nat Rev Endocrionol 2017;13(3):161-173.

(17) Gore AC, Martien KM, Gagnidze K, Pfaff D. Implications of prenatal steroid perturbations for neurodevelopment, behavior, and autism. Endocr Rev 2014;35(6):961-991.

(18) Rivollier F, Krebs M-O, Kebir O. Perinatal exposure to environmental endocrine disruptors in the emergence of neurodevelopmental psychiatric diseases: A systematic review. Int J Environ Res Public Health 2019;16(8):9318.

(19) Schug TT, Blawas AM, Gray K, Heindel JJ, Lawler CP. Elucidating the links between endocrine disruptors and neurodevelopment. Endocrinology 2015;156(6):1941-1951.

(20) Gogolla N, LeBlanc JJ, Quast KB, Sudhof TC, Fagiolini M, Hensch TK. Common circuit defect of excitatory-inhibitory balance in mouse models of autism. JOURNAL OF NEURODEVELOPMENTAL DISORDERS 2009;1(2):172-181.

(21) Lichtensteiger W, Bassetti-Gaille C, Faass O, Axelstad M, Boberg J, Christiansen S, et al. Differential gene expression patterns in developing sexually dimorphic rat brain regions exposed to antiandrogenic, estrogenic, or complex endocrine disruptor mixtures: glutamatergic synapses as target. Endocrinology 2015;156(4):1477-1493.

(22) Wolstenholme JJ, Goldsby JA, Rissman EF. Transgenerational effects of prenatal bisphenol A on social recognition. Horm Behav 2013;64(5):833-39.

(23) Sethi S, Keil KP, Chen H, Hayakawa K, Li X, Lin Y, et al. Detection of 3,3'-Dichlorobiphenyl in human maternal plasma and its effects on axonal and dendritic growth in primary rat neurons. Toxicol Sci 2017;158(2):401-411.

(24) Eskenazi B, Rauch SA, Tenerelli R, Huen K, Holland NT, Lustig RH, et al. In utero and childhood DDT, DDE, PBDE and PCBs exposure and sex hormones in adolescent boys: The CHAMACOS study. Int J Hyg Environ Health 2017;220(2):364-372.

(25) Manikkam M, Guerrero-Bosagna C, Tracey R, Haque MM, Skinner MK. Transgenerational actions of environmental compounds on reproductive disease and identification of epigenetic biomarkers of ancestral exposures. PLoS ONE 2012;7(2):e31901.

(26) McBirney M, King SE, Pappalardo M, Houser E, Unkefer M, Nilsson E, et al. Atrazine induced epigenetic transgenerational inheritance of disease, lean phenotype and sperm epimutation pathology biomarkers. PLoS ONE 2017;12:e0184306.

(27) Dunaway KW, Islam MS, Coulson RL, Lopez SJ, Vogel Ciernia A, Chu RG, et al. Cumulative Impact of Polychlorinated Biphenyl and Large Chromosomal Duplications on DNA Methylation, Chromatin, and Expression of Autism Candidate Genes. Cell Rep 2016;17(11):3035-3048.

(28) Mitchell MM, Woods R, Chi L-H, Schmidt RJ, Pessah IN, Kostyniak PJ, et al. Levels of select PCB and PBDE congeners in human postmortem brain reveal possible environmental involvement in 15q11-q13 duplication autism spectrum disorder. Environ Mol Mutagen 2012;53(8):589-598.

(29) Bakulski KM, Singer AB, Fallin MD. Genes and environment in autism spectrum disorders: An integrated perspective. In: Hu VW, editor. Frontiers in Autism Research: New Horizons for Diagnosis and Treatment. Singapore: World Scientific Pubishing Co.; 2014. p. 335-374.

(30) Lasalle JM. Epigenomic strategies at the interface of genetic and environmental risk factors for autism. J Hum Genet 2013;58(7):396-401.

(31) Moosa A, Shu H, Sarachana T, Hu VW. Are endocrine disrupting compounds environmental risk factors for autism spectrum disorder? Horm Behav 2018;101:13-21.

(32) Tran NQV, Miyake K. Neurodevelopmental Disorders and Environmental Toxicants: Epigenetics as an Underlying Mechanism. Int J Genomics 2017;2017:7526592.

(33) De Coster S, Van Larebeke N. Endocrine-disrupting chemicals: Associated disorders and mechanisms of action. J Environ Public Health 2012;2012:713696.

(34) Rignell-Hydbom A, Rylander L, Giwercman A, Jönsson BAG, Lindh C, Eleuteri P, et al. Exposure to PCBs and p,p′-DDE and human sperm chromatin integrity. Environ Health Perspect 2005;113(2):175-179.

(35) McAuliffe ME, Williams PL, Korrick SA, Altshul LM, Perry MJ. Environmental exposure to polychlorinated biphenyls and p,p ´-DDE and sperm sex-chromosome disomy. Environ Health Perspect 2012;120(4):535-540.

(36) Perry MJ, Young HA, Grandjean P, Halling J, Petersen MS, Martenies SE, et al. Sperm aneuploidy in faroese men with lifetime exposure to dichlorodiphenyldichloroethylenchandigarhe(P,p´-DDE) and polychlorinated biphenyl (PCB) pollutants. Environ Health Perspect 2016;124(7):951-956.

(37) Wu H, Hauser R, Krawetz SA, Pilsner JR. Environmental Susceptibility of the Sperm Epigenome During Windows of Male Germ Cell Development. Curr Environ Health Rep 2015;2(4):356-366.

(38) Jenkins TG, Aston KI, Pflueger C, Cairns BR, Carrell DT. Age-Associated Sperm DNA Methylation Alterations: Possible Implications in Offspring Disease Susceptibility. PLoS Genetics 2014;10(7):e1004458.

(39) Milekic MH, Xin Y, O'Donnell A, Kumar KK, Bradley-Moore M, Malaspina D, et al. Age-related sperm DNA methylation changes are transmitted to offspring and associated with abnormal behavior and dysregulated gene expression. Mol Psychiatry 2015;20(8):995-1001.

(40) Quinlan CA, McVeigh KH, Driver CR, Govind P, Karpati A. Parental Age and Autism Spectrum Disorders Among New York City Children 0–36 Months of Age. Matern Child Health J 2015;19(8):1783-90.

(41) Reichenberg A, Gross R, Weiser M, Bresnahan M, Silverman J, Harlap S, et al. Advancing paternal age and autism. Arch Gen Psychiatry 2006;63(9):1026-32.

(42) Vierck E, Silverman JM. Brief Report: Phenotypic Differences and their Relationship to Paternal Age and Gender in Autism Spectrum Disorder. J Autism Dev Disord 2014;45(6):1915-1924.

(43) Feinberg JI, 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. International Journal of Epidemiology 2015;14:1-12.

(44) Garrido N, Cruz F, Egea RR, Simon C, Sadler-Riggleman I, Beck D, et al. Sperm DNA methylation epimutation biomarker for paternal offspring autism susceptibility. Clin Epigenetics 2021;13(1):6.

(45) Grandjean P, Weihe P, Burse VW, Needham LL, Storr-Hansen E, Heinzow B, et al. Neurobehavioral deficits associated with PCB in 7-year-old children prenatally exposed to seafood neurotoxicants. Neurotoxicol Teratol 2001;23(4):305-317.

(46) Wu H, de Gannes MK, Luchetti G, Richard Pilsner J. Rapid method for the isolation of mammalian sperm DNA. BioTechniques 2015;58(6):293-300.

(47) Krueger F, Andrews SR. Bismark: A flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics 2011;27(11):1571-1572.

(48) 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):672-684.

(49) Hansen KD, Langmead B, Irizarry RA. BSmooth: from whole genome bisulfite sequencing reads to differentially methylated regions. Genome Biol 2012;13(10)R83.

(50) Korthauer K, Chakraborty S, Benjamini Y, Irizarry RA. Detection and accurate false discovery rate control of differentially methylated regions from whole genome bisulfite sequencing. Biostatistics 2019;20(3):367-383.

(51) Cavalcante RG, Sartor MA. Annotatr: Genomic regions in context. Bioinformatics 2017;33(15):2381-2383.

(52) Gu Z. rGREAT: Client for GREAT Analysis. R package version 1.0.0. 2014. https://github.com/jokergoo/rGREAT: Bioconductor version: Release 3.1.

(53) Grote S. GOfuncR: Gene ontology enrichment using FUNC. R package version 1.8.0. 2020. https://bioconductor.riken.jp/packages/release/bioc/html/GOfuncR.html: Bioconductor version: Release 3.11.

(54) Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res 2019;47(D1):D607-D613.

(55) Supek F, Bošnjak M, Škunca N, Šmuc T. Revigo summarizes and visualizes long lists of gene ontology terms. PLoS ONE 2011;6(7):e21800.

(56) Degenhardt F, Seifert S, Szymczak S. Evaluation of variable selection methods for random forests and omics data sets. Brief Bioinform 2019;20(2):492-503.

(57) Kursa MB, Rudnicki WR. Feature selection with the boruta package. J Stat Software 2010;36(11):1-13.

(58) Das P, Roychowdhury A, Das S, Roychoudhury S, Tripathy S. sigFeature: Novel Significant Feature Selection Method for Classification of Gene Expression Data Using Support Vector Machine and t Statistic. Front Genet 2020;11:247.

(59) Abrahams BS, Arking DE, Campbell DB, Mefford HC, Morrow EM, Weiss LA, et al. SFARI Gene 2.0: A community-driven knowledgebase for the autism spectrum disorders (ASDs). Mol Autism 2013;4(1):36.

(60) Oliveros JC. Venny. An interactive tool for comparing lists with Venn's diagrams. 2007. http://bioinfogp.cnb.csic.es/tools/venny/index.html

(61) Leung Y-K, Ouyang B, Niu L, Xie C, Ying J, Medvedovic M, et al. Identification of sex-specific DNA methylation changes driven by specific chemicals in cord blood in a Faroese birth cohort. Epigenetics 2018;13(3):290-300.

(62) 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. Genome Med 2020;12(1):88.

(63) 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.

(64) Hu VW, Hong Y, Xu M, Shu HT. Altered DNA methylation in a severe subtype of idiopathic autism: Evidence for sex differences in affected metabolic pathways. Autism : the international journal of research and practice 2020; doi:10.1177/1362361320971085.

(65) Schroeder DI, Schmidt RJ, Crary-Dooley FK, Walker CK, Ozonoff S, Tancredi DJ, et al. Placental methylome analysis from a prospective autism study. Mol Autism 2016;7(1):51.

(66) Su J, Huang Y-H, Cui X, Wang X, Zhang X, Lei Y, et al. Homeobox oncogene activation by pan-cancer DNA hypermethylation. Genome Biol 2018;19(1):108.

(67) Dorgan JF, Brock JW, Rothman N, Needham LL, Miller R, Stephenson Jr. HE, et al. Serum organochlorine pesticides and PCBs and breast cancer risk: Results from a prospective analysis (USA). Cancer Causes Control 1999;10(1):1-11.

(68) Jaga K, Dharmani C. The epidemiology of pesticide exposure and cancer: A review. Rev Environ Health 2005;20(1):15-38.

(69) Purdue MP, Hoppin JA, Blair A, Dosemeci M, Alavanja MCR. Occupational exposure to organochlorine insecticides and cancer incidence in the Agricultural Health Study. Int J Cancer 2007;120(3):642-649.

(70) Crawley JN, Heyer W-D, LaSalle JM. Autism and Cancer Share Risk Genes, Pathways, and Drug Targets. Trends Genet 2016;32(3):139-146.

(71) Molenaar JJ, Koster J, Zwijnenburg DA, Van Sluis P, Valentijn LJ, Van Der Ploeg I, et al. Sequencing of neuroblastoma identifies chromothripsis and defects in neuritogenesis genes. Nature 2012;483(7391):589-593.

(72) Cukier HN, Dueker ND, Slifer SH, Lee JM, Whitehead PL, Lalanne E, et al. Exome sequencing of extended families with autism reveals genes shared across neurodevelopmental and neuropsychiatric disorders. Mol Autism 2014;5(1):1.

(73) Guo H, Peng Y, Hu Z, Li Y, Xun G, Ou J, et al. Genome-wide copy number variation analysis in a Chinese autism spectrum disorder cohort. Sci Rep 2017;7:44155.

(74) Hu VW, Addington A, Hyman A. Novel autism subtype-dependent genetic variants are revealed by quantitative trait and subphenotype association analyses of Published GWAS Data. PLoS ONE 2011;6(4):e19067.

(75) Woo HJ, Yu C, Kumar K, Reifman J. Large-scale interaction effects reveal missing heritability in schizophrenia, bipolar disorder and posttraumatic stress disorder. Transl Psychiatry 2017;7(4):e1089.

(76) Gauthier J, Siddiqui TJ, Huashan P, Yokomaku D, Hamdan FF, Champagne N, et al. Truncating mutations in NRXN2 and NRXN1 in autism spectrum disorders and schizophrenia. Hum Genet 2011;130(4):563-573.

(77) Mohrmann I, Gillessen-Kaesbach G, Siebert R, Caliebe A, Hellenbroich Y. A de novo 0.57 Mb microdeletion in chromosome 11q13.1 in a patient with speech problems, autistic traits, dysmorphic features and multiple endocrine neoplasia type 1. Eur J Med Genet 2011;54(4):e461-e464.

(78) Dachtler J, Ivorra JL, Rowland TE, Lever C, John Rodgers R, Clapcote SJ. Heterozygous deletion of α-neurexin I or α-neurexin II results in behaviors relevant to autism and schizophrenia. Behav Neurosci 2015;129(6):765-776.

(79) Harkin LF, Lindsay SJ, Xu Y, Alzu'Bi A, Ferrara A, Gullon EA, et al. Neurexins 1-3 each have a distinct pattern of expression in the early developing human cerebral cortex. Cereb Cortex 2017;27(1):216-232.

(80) Bacchelli E, Cameli C, Viggiano M, Igliozzi R, Mancini A, Tancredi R, et al. An integrated analysis of rare CNV and exome variation in Autism Spectrum Disorder using the Infinium PsychArray. Sci Rep 2020;10(1):3198.

(81) Griswold AJ, Dueker ND, Van Booven D, Rantus JA, Jaworski JM, Slifer SH, et al. Targeted massively parallel sequencing of autism spectrum disorder-associated genes in a case control cohort reveals rare loss-of-function risk variants. Mol Autism 2015;6(1):43.

(82) Martin CL, Duvall JA, Ilkin Y, Simon JS, Arreaza MG, Wilkes K, et al. Cytogenetic and molecular characterization of A2BP1/FOX1 as a candidate gene for autism. Am J Med Genet B Neuropsychiatr Genet 2007;144B(7):869-76.

(83) Lin X, Lim IY, Wu Y, Teh AL, Chen L, Aris IM, et al. Developmental pathways to adiposity begin before birth and are influenced by genotype, prenatal environment and epigenome. BMC Med 2017;15(1):50.

(84) Butler AE, Hayat S, Dargham SR, Malek JA, Abdulla SA, Mohamoud YA, et al. Alterations in long noncoding RNAs in women with and without polycystic ovarian syndrome. Clin Endocrinol 2019;91(6):793-797.

(85) Galardi S, Fatica A, Bachi A, Scaloni A, Presutti C, Bozzoni I. Purified box C/D snoRNPs are able to reproduce site-specific 2′-O-methylation of target RNA in vitro. Mol Cell Biol 2002;22(19):6663-6668.

(86) Bratkovič T, Modic M, Camargo Ortega G, Drukker M, Rogelj B. Neuronal differentiation induces SNORD115 expression and is accompanied by post-transcriptional changes of serotonin receptor 2c mRNA. Sci Rep 2018;8(1):5101.

(87) Cavaillé J. Box C/D small nucleolar RNA genes and the Prader-Willi syndrome: a complex interplay. Wiley Interdiscip Rev RNA 2017;8(4). doi:10.1002/wrna.1417.

(88) Raabe CA, Voss R, Kummerfeld D-, Brosius J, Galiveti CR, Wolters A, et al. Ectopic expression of Snord115 in choroid plexus interferes with editing but not splicing of 5-Ht2c receptor pre-mRNA in mice. Sci Rep 2019;9(1):4300.

(89) Faulk C, Kim JH, Jones TR, McEachin RC, Nahar MS, Dolinoy DC, et al. Bisphenol A-associated alterations in genome-wide DNA methylation and gene expression patterns reveal sequence-dependent and non-monotonic effects in human fetal liver. Environmental Epigenetics 2015;1(1):dvv006.

(90) Bürger J, Horn D, Tönnies H, Neitzel H, Reis A. Familial interstitial 570 kbp deletion of the UBE3A gene region causing Angelman syndrome but not Prader-Willi syndrome. Am J Med Genet 2002;111(3):233-237.

(91) Runte M, Varon R, Horn D, Horsthemke B, Buiting K. Exclusion of the C/D box snoRNA gene cluster HBII-52 from a major role in Prader-Willi syndrome. Hum Genet 2005;116(3):228-230.

(92) Depienne C, Moreno-De-Luca D, Heron D, Bouteiller D, Gennetier A, Delorme R, et al. Screening for Genomic Rearrangements and Methylation Abnormalities of the 15q11-q13 Region in Autism Spectrum Disorders. Biol Psychiatry 2009;66(4):349-359.

(93) Finucane BM, Lusk L, Arkilo D, Chamberlain S, Devinsky O, Dindot S, et al. 15q duplication syndrome and related disorders. In: Adam MP, Ardinger HH, Pagon RA, Wallace SE, Bean LJH, Stephens K, et al, editors. GeneReviews (Internet). 2016 ed. Internet: University of Washington, Seattle, WA; 2016. https://www.ncbi.nlm.nih.gov/books/NBK367946/

(94) Scoles HA, Urraca N, Chadwick SW, Reiter LT, Lasalle JM. Increased copy number for methylated maternal 15q duplications leads to changes in gene and protein expression in human cortical samples. Molecular Autism 2011;2(1):19.

(95) Anway MD, Skinner MK. Epigenetic programming of the germ line: Effects of endocrine disruptors on the development of transgenerational disease. Reproductive BioMedicine Online 2008;16(1):23-25.

(96) Skinner MK, Manikkam M, Guerrero-Bosagna C. Epigenetic transgenerational actions of endocrine disruptors. Reproductive Toxicology 2011;31(3):337-343.

(97) Anway MD, Cupp AS, Uzumcu N, Skinner MK. Toxicology: Epigenetic transgenerational actions of endocrine disruptors and male fertility. Science 2005;308(5727):1466-1469.

(98) Manikkam M, Tracey R, Guerrero-Bosagna C, Skinner MK. Plastics Derived Endocrine Disruptors (BPA, DEHP and DBP) Induce Epigenetic Transgenerational Inheritance of Obesity, Reproductive Disease and Sperm Epimutations. PLoS ONE 2013;8(1):e55387.

(99) Vandenberg LN, Colborn T, Hayes TB, Heindel JJ, Jacobs DR Jr, Lee D, et al. Hormones and endocrine-disrupting chemicals: low-dose effects and nonmonotonic dose responses. Endocr Rev 2012;33(3):378-455.

Figure legends

Figure 1. Overview of workflow for this study

Figure 2. Heatmaps (A, B) and Manhattan plots (C, D) depicting the results of the WGBS analyses.
DMRs from the Discovery set (A) and Validation set (B) are shown in the heatmaps. All significant DMRs are represented above the horizontal lines in the Manhattan plots.

Figure 3. *SNORD115* region showing significant genome-wide differential methylation between exposure groups after correction for multiple testing

Figure 4. Significant gene ontology terms from GofuncR analyses of DMRs from discovery and validation cohorts.
Specific p-values are from a meta p-value analysis of the least dispensable significant ($p < 0.05$) gene ontology terms.

Figure 5. Overlap of DDE DMR-associated genes among the discovery and validation sets and SFARI genes

Figure 6. Results of pyrosequencing analyses of DMRs associated with *CSMD1* (A, B), *RBFOX1* (C), and *NRXN2* (D).
The box plot (A) shows differential methylation at a single CpG site in *CSMD1* while the graphs show the average methylation as a function of DDE serum concentration (μg/gm lipid) for *CSMD1*, *RBFOX1* (10 sites, discovery set only), and *NRXN2* (7 sites, all samples in both discovery and validation sets). R-squared (r^2) and p-values for the correlation curves are shown.

Figure 7. Results of pyrosequencing analyses of DMRs associated with *MIRLET7BHG.*
The graph shows the average methylation across 4 CpG sites as a function of DDE serum concentration (μg/gm lipid) while the table below shows the methylation differences at each site.

List of Additional Files

Additional file 1. Summary of demographic and EDC data for sperm donors in each exposure tertile

Additional file 2. Correlations between serum concentrations of DDE and either DDT or Σ(4 prevalent PCB congeners)

Additional file 3. Relationship between serum concentrations of DDE and smoking status or BMI

Additional file 4. Correlations between serum concentrations of DDE and sperm parameters

Additional file 5. DMRs identified in the discovery set

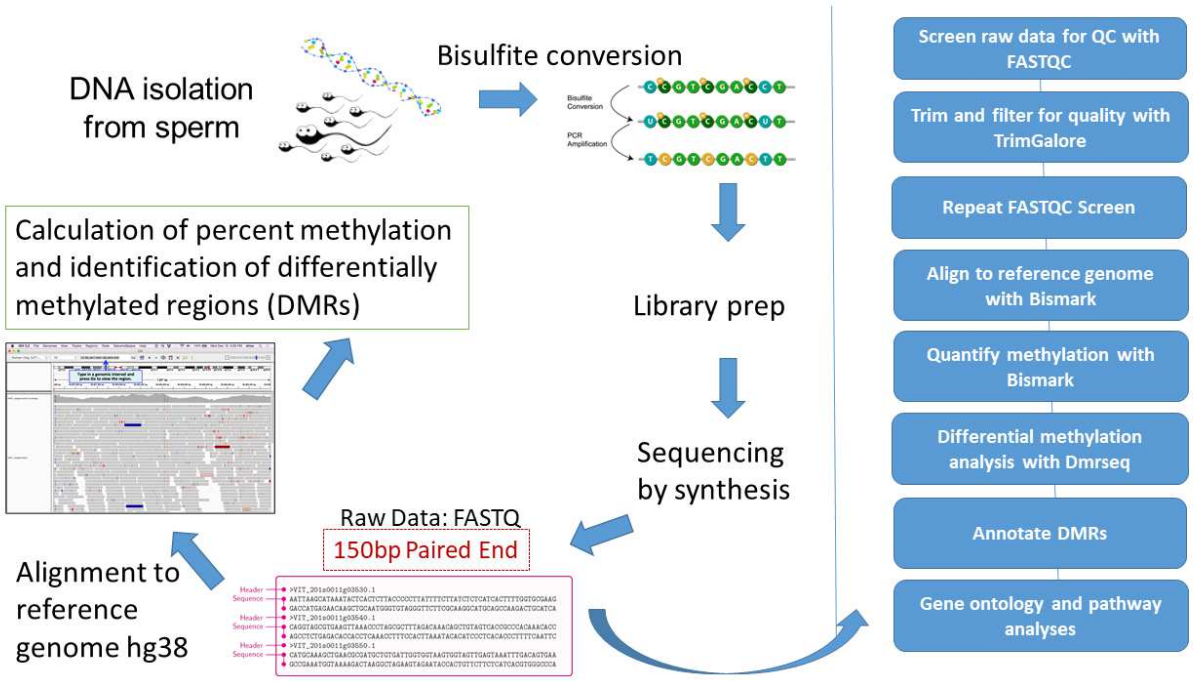
Additional file 6. DMRs identified in the validation set

Additional file 7: Primer sequences and PCR conditions for pyrosequencing analyses

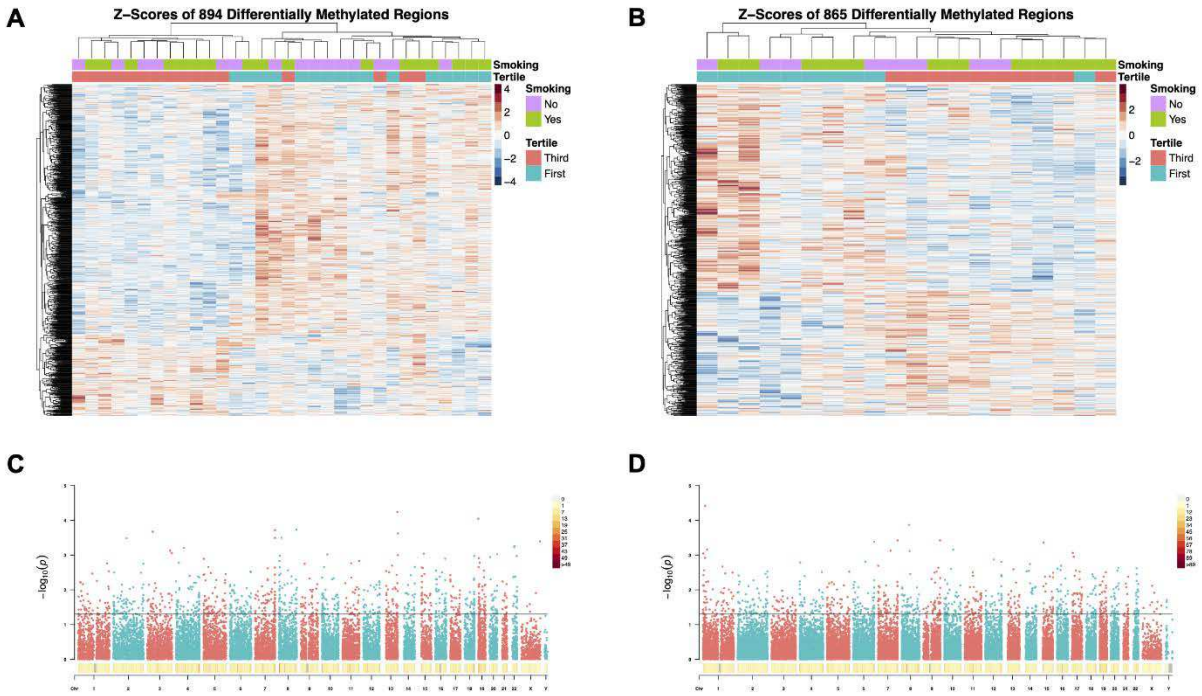
Additional file 8. Annotated genomic sites/regions among DMRs in discovery and validation sets

Additional file 9. DMRs identified in the combined set of 52 samples

Figure 1.


DNA isolation from sperm
Bisulfite conversion
Library prep
Sequencing by synthesis
Raw Data: FASTQ
150bp Paired End
Alignment to reference genome hg38
Calculation of percent methylation and identification of differentially methylated regions (DMRs)
Screen raw data for QC with FASTQC
Trim and filter for quality with TrimGalore
Repeat FASTQC Screen
Align to reference genome with Bismark
Quantify methylation with Bismark
Differential methylation analysis with Dmrseq
Annotate DMRs
Gene ontology and pathway analyses

990 Figure 2.
991



992
993
994

Figure 3.

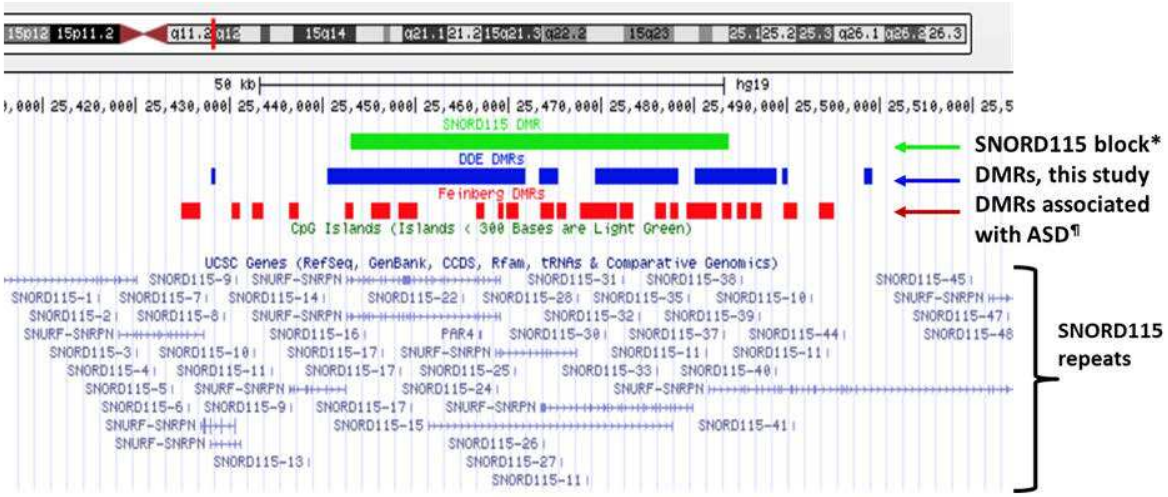

SNORD115 block*
DMRs, this study
DMRs associated with ASD¶
SNORD115 repeats
SNORD115 DMR
DOE DMRs
Feinberg DMRs
CpG Islands (Islands < 300 Bases are Light Green)
UCSC Genes (RefSeq, GenBank, CCDS, Rfam, tRNAs & Comparative Genomics)

Figure 4.

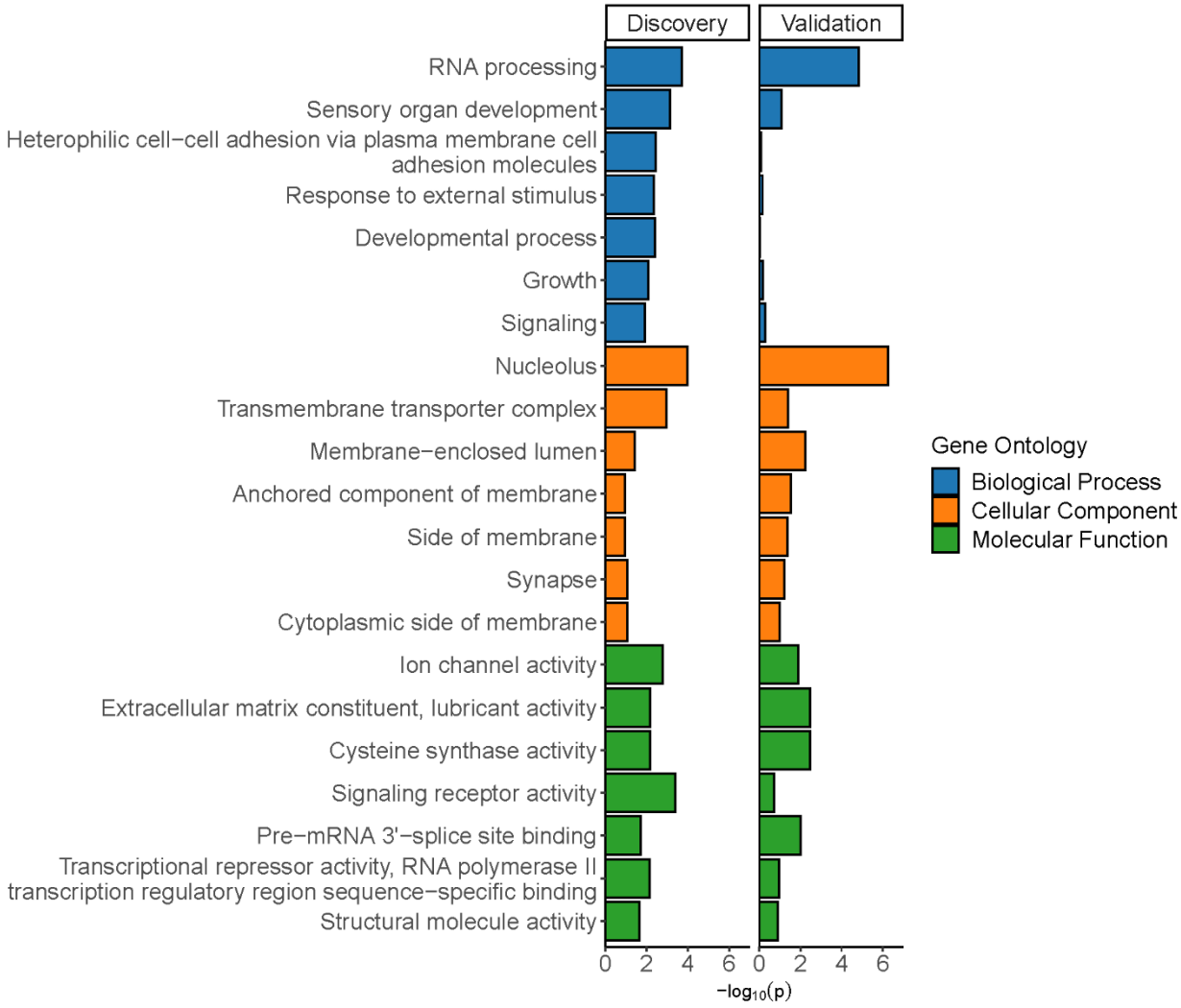
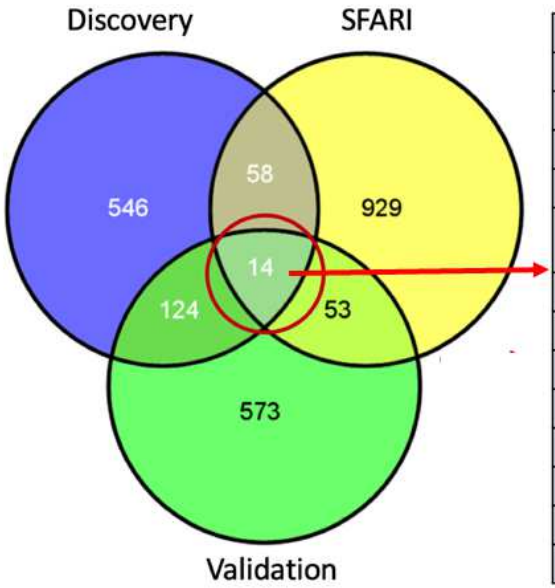

Discovery
Validation
RNA processing
Sensory organ development
Heterophilic cell−cell adhesion via plasma membrane cell adhesion molecules
Response to external stimulus
Developmental process
Growth
Signaling
Nucleolus
Transmembrane transporter complex
Membrane−enclosed lumen
Anchored component of membrane
Side of membrane
Synapse
Cytoplasmic side of membrane
Ion channel activity
Extracellular matrix constituent, lubricant activity
Cysteine synthase activity
Signaling receptor activity
Pre−mRNA 3'−splice site binding
Transcriptional repressor activity, RNA polymerase II transcription regulatory region sequence−specific binding
Structural molecule activity
0 2 4 6
0 2 4 6
−log10(p)
Gene Ontology
Biological Process
Cellular Component
Molecular Function

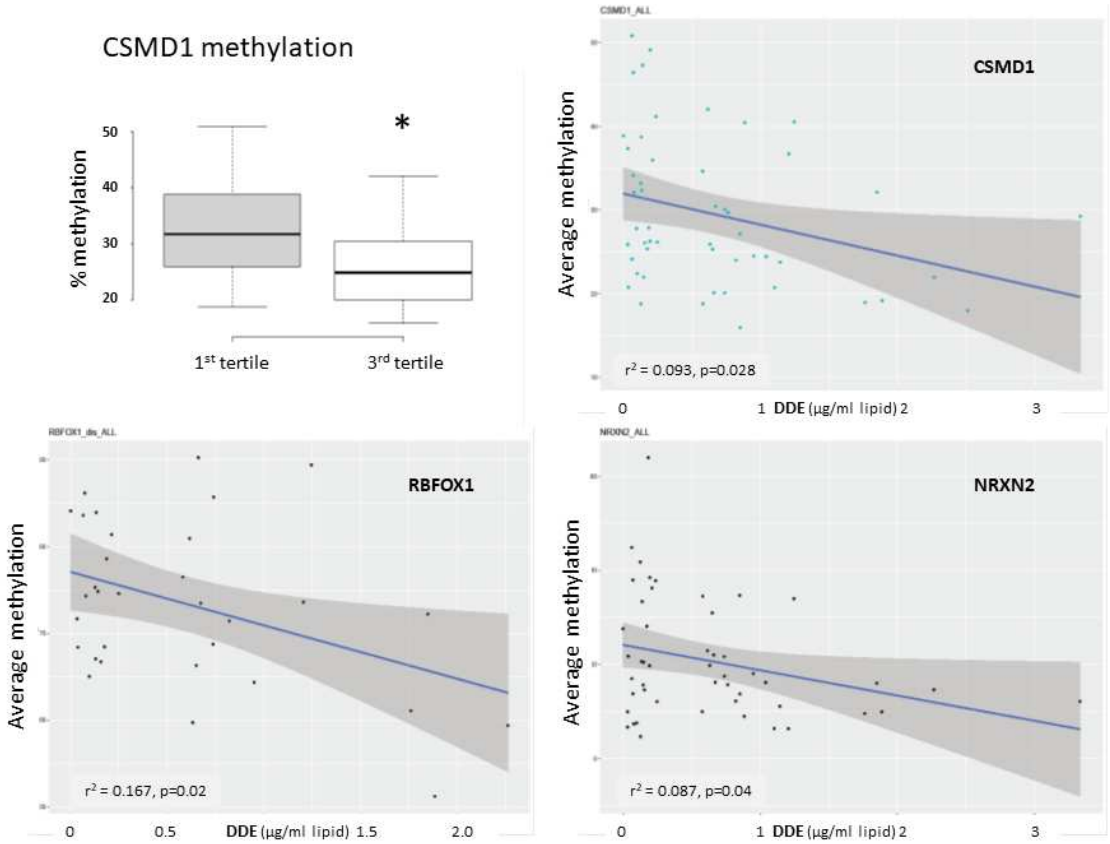
Figure 5.



CSMD1	CUB and sushi domain-containing protein 1
HDAC4	Histone deacetylase 4
RBFOX1	RNA binding protein fox-1 homolog 1
CHRM3	Muscarinic acetylcholine receptor M3
JARID2	Protein Jumonji
KHDRBS2	KH domain-containing, RNA-binding, signal transduction-associated protein 2
TM4SF20	Transmembrane 4 L6 family member 20
EBF3	Transcription factor COE3
GRM5	Metabotropic glutamate receptor 5
CDH9	Cadherin-9
NRXN2	Neurexin-2
TSHZ3	Teashirt homolog 3
SYT1	Synaptotagmin-1
CTNNB1	Catenin beta-1

DEVLOPMENT

1012 Figure 6.
1013
1014

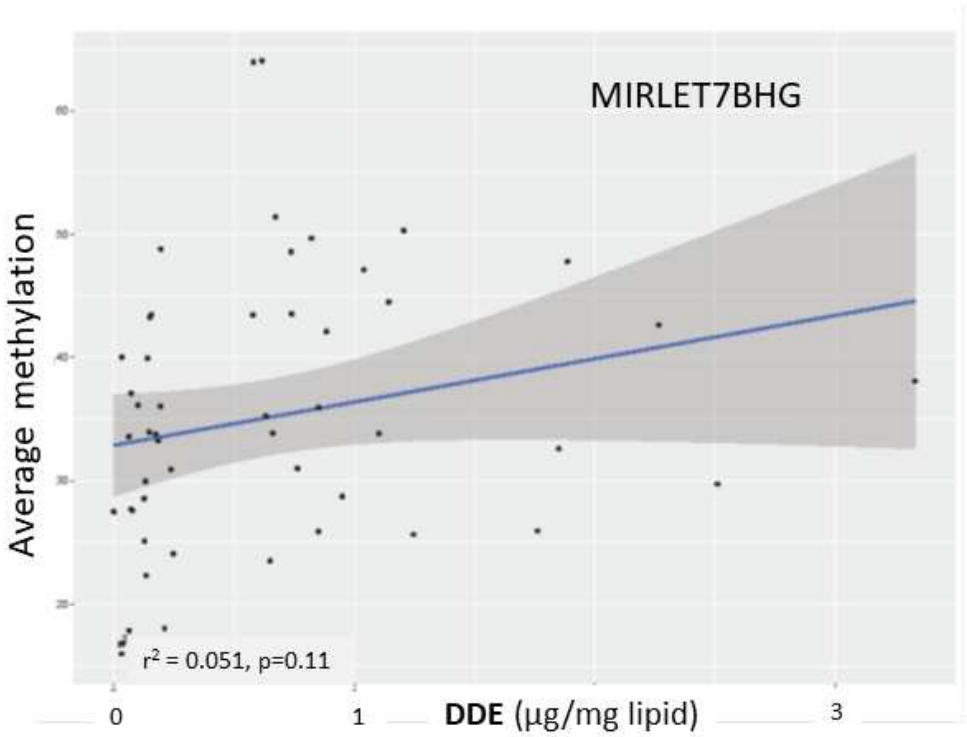


1015
1016

1017 Figure 7.

1018

1019



	MIRLET7BHG*				Average
Tertile	Pos. 1	Pos. 2	Pos. 3	Pos. 4	All positions
First	15.03	14.31	36.41	55.47	30.30
Third	21.82	23.32	47.50	67.20	39.96
Difference (Third-First)	6.79	9.01	11.09	11.73	9.66

*Methylation was quantified at 4 specific positions in the MIRLET7BHG DMR

1020

Figures

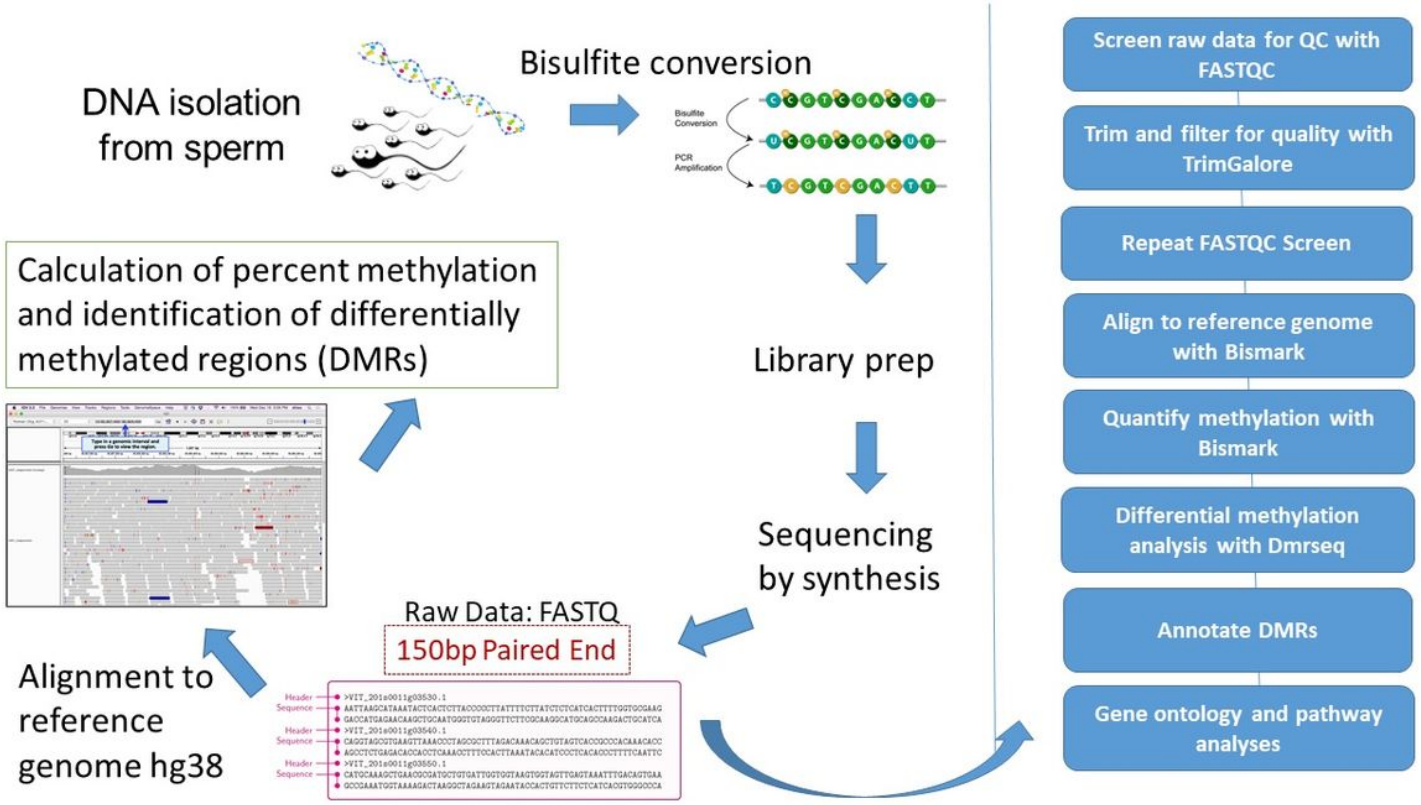


Figure 1

Overview of workflow for this study

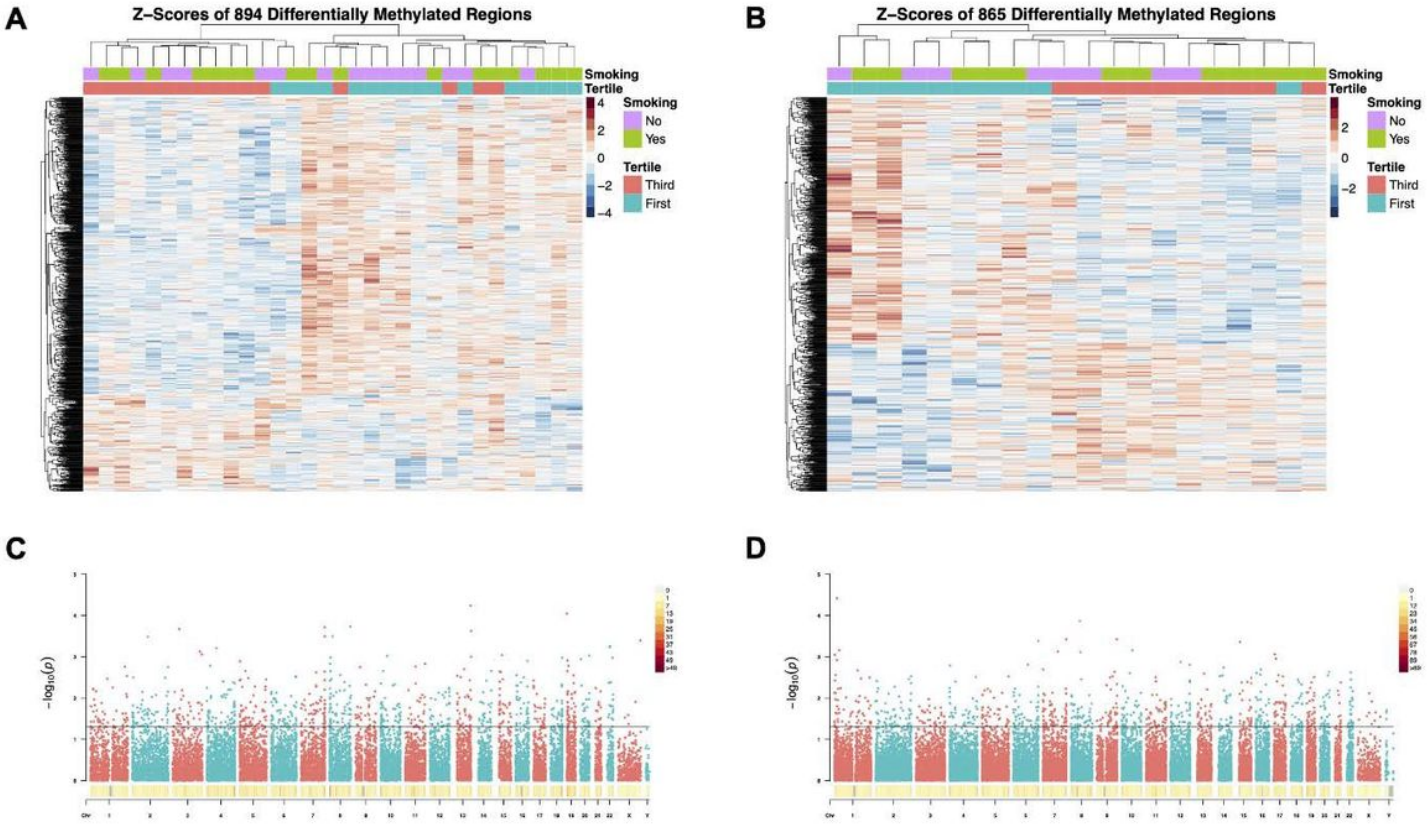


Figure 2

Heatmaps (A, B) and Manhattan plots (C, D) depicting the results of the WGBS analyses. DMRs from the Discovery set (A) and Validation set (B) are shown in the heatmaps. All significant DMRs are represented above the horizontal lines in the Manhattan plots.

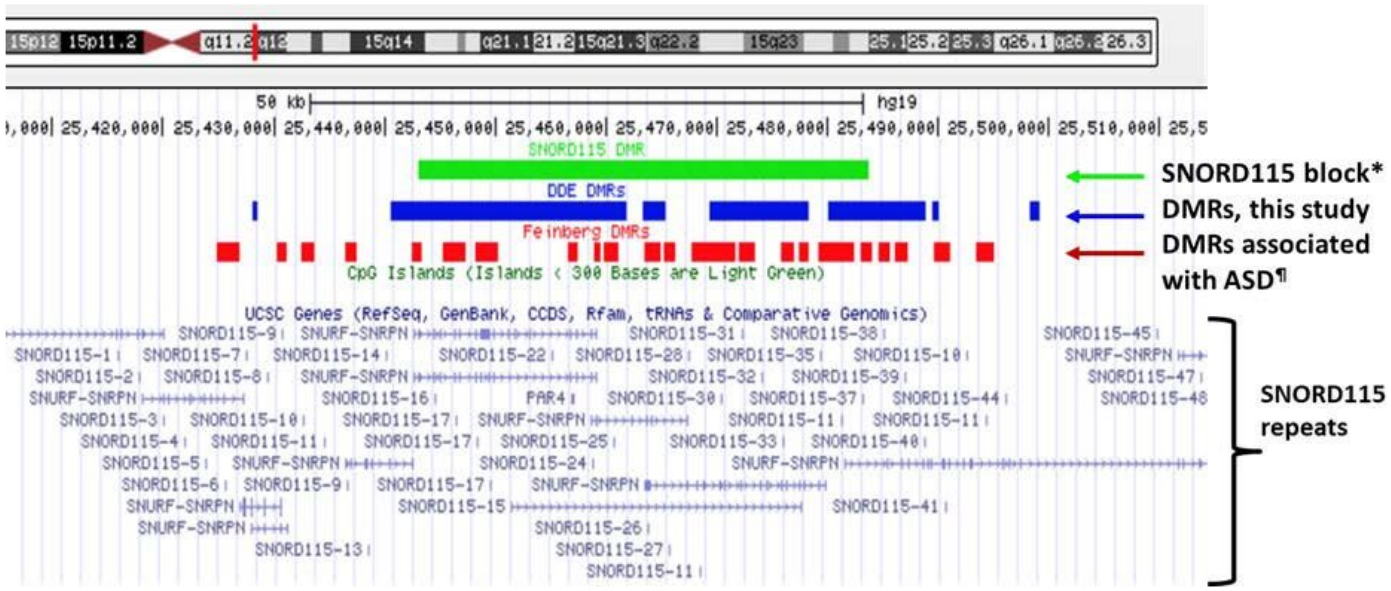


Figure 3

SNORD115 region showing significant genome-wide differential methylation between exposure groups after correction for multiple testing

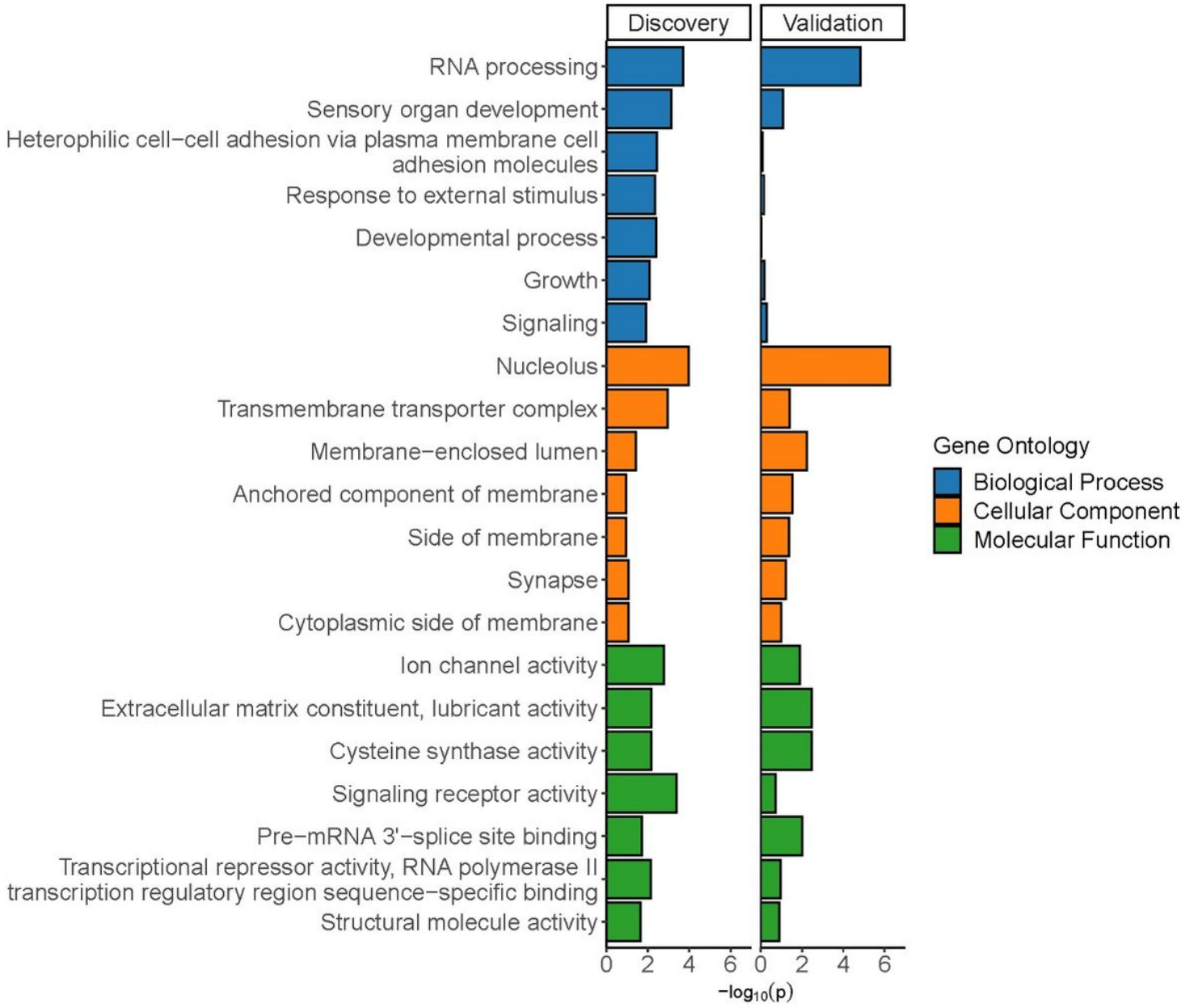
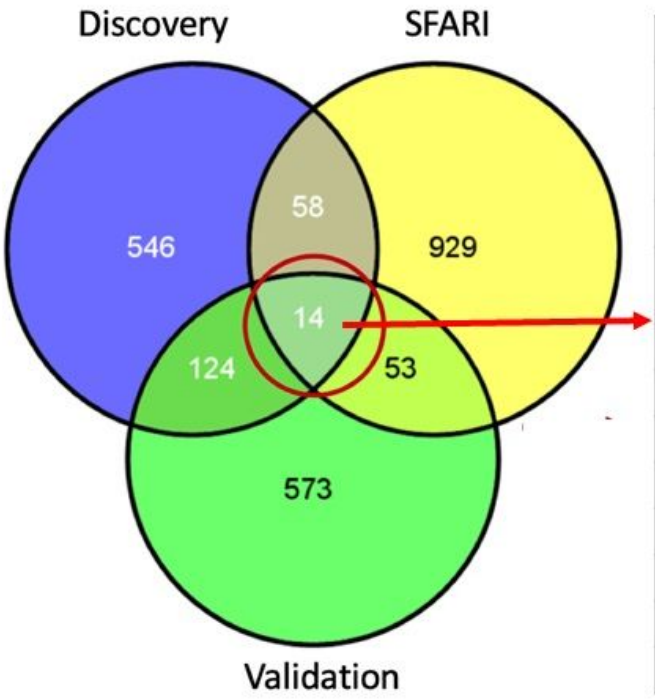


Figure 4

Significant gene ontology terms from GofuncR analyses of DMRs from discovery and validation cohorts. Specific p-values are from a meta p-value analysis of the least dispensable significant ($p < 0.05$) gene ontology terms.



CSMD1	CUB and sushi domain-containing protein 1
HDAC4	Histone deacetylase 4
RBFOX1	RNA binding protein fox-1 homolog 1
CHRM3	Muscarinic acetylcholine receptor M3
JARID2	Protein Jumonji
KHDRBS2	KH domain-containing, RNA-binding, signal transduction-associated protein 2
TM4SF20	Transmembrane 4 L6 family member 20
EBF3	Transcription factor COE3
GRM5	Metabotropic glutamate receptor 5
CDH9	Cadherin-9
NRXN2	Neurexin-2
TSHZ3	Teashirt homolog 3
SYT1	Synaptotagmin-1
CTNNB1	Catenin beta-1

DEVLOPMENT

Figure 5

Overlap of DDE DMR-associated genes among the discovery and validation sets and SFARI genes

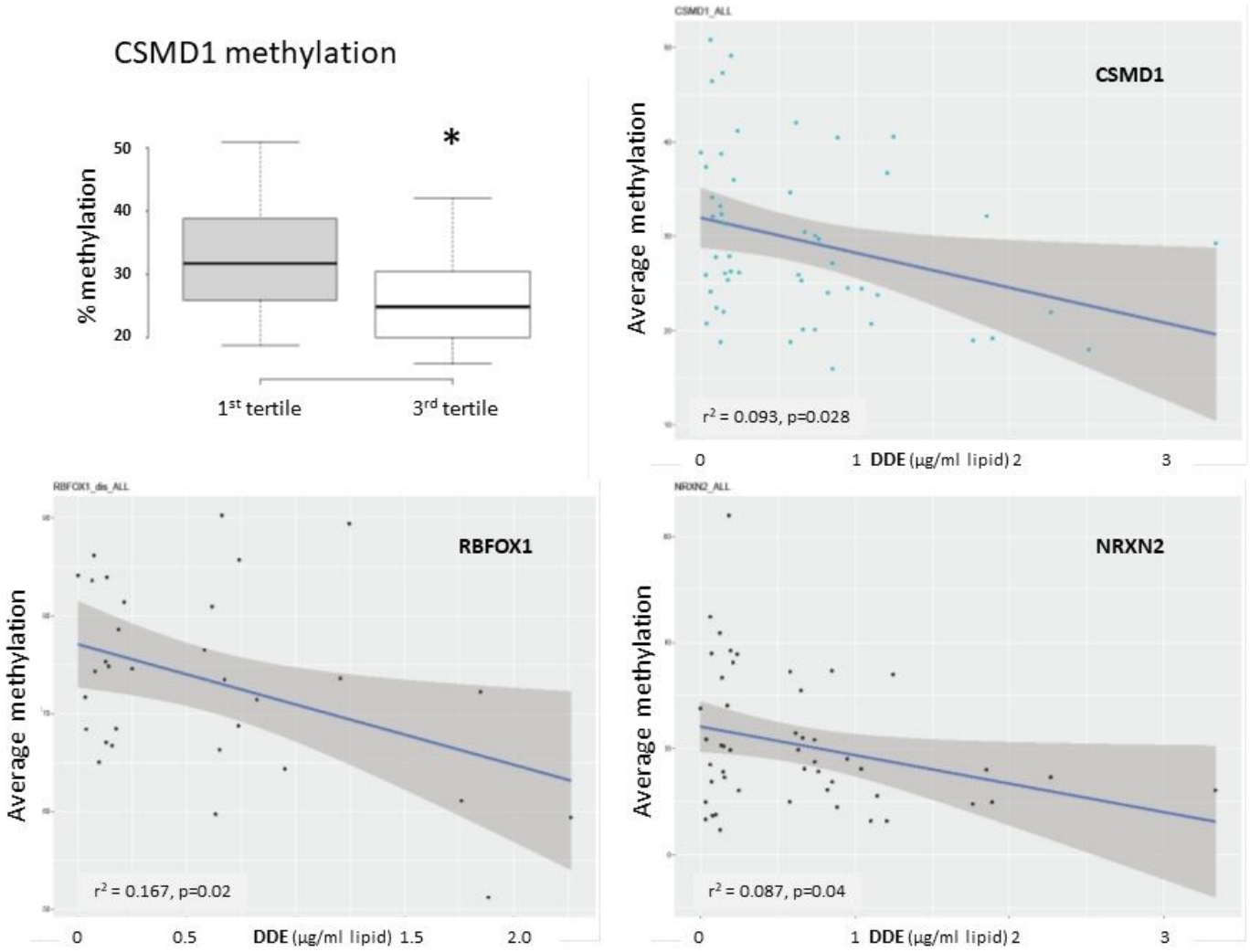
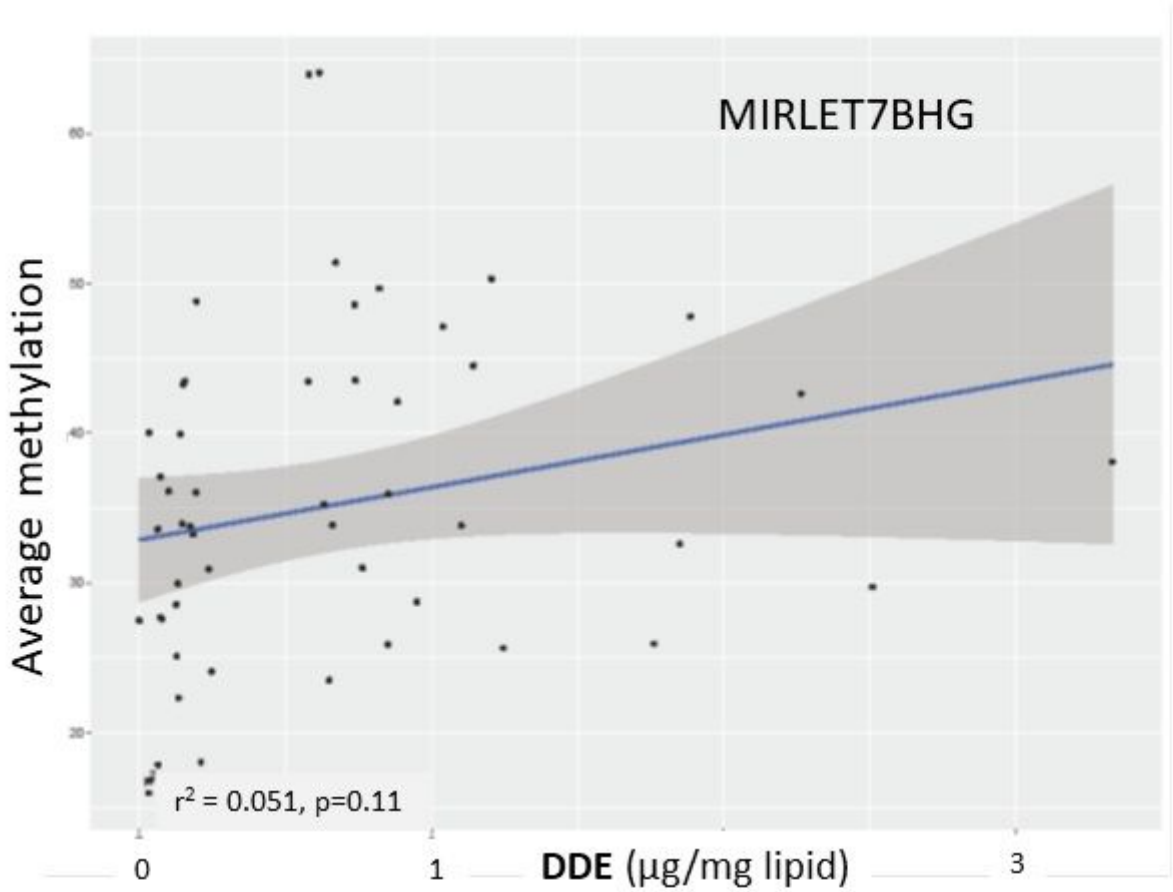


Figure 6

Results of pyrosequencing analyses of DMRs associated with CSMD1 (A, B), RBFOX1 (C), and NRXN2 (D). The box plot (A) shows differential methylation at a single CpG site in CSMD1 while the graphs show the average methylation as a function of DDE serum concentration (µg/gm lipid) for CSMD1, RBFOX1 (10 sites, discovery set only), and NRXN2 (7 sites, all samples in both discovery and validation sets). R-squared (r2) and p-values for the correlation curves are shown.



	MIRLET7BHG*				Average
Tertile	Pos. 1	Pos. 2	Pos. 3	Pos. 4	All positions
First	15.03	14.31	36.41	55.47	30.30
Third	21.82	23.32	47.50	67.20	39.96
Difference (Third-First)	6.79	9.01	11.09	11.73	9.66

*Methylation was quantified at 4 specific positions in the MIRLET7BHG DMR

Figure 7

Results of pyrosequencing analyses of DMRs associated with MIRLET7BHG. The graph shows the average methylation across 4 CpG sites as a function of DDE serum concentration (µg/gm lipid) while the table below shows the methylation differences at each site.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- AdditionalfilesMolecularAutism.7z