Title

Exposure to air pollution is associated with DNA methylation changes in sperm.

Journal

Environmental Epigenetics, 10(1)

Authors

Schrott, Rose

Feinberg, Jason

Newschaffer, Craig

et al.

Publication Date

2024

DOI

10.1093/eep/dvae003

Peer reviewed

ENVIRONMENTAL EPIGENETICS

Environmental Epigenetics, 2024, **10(1)**, dvae003

DOI: https://doi.org/10.1093/eep/dvae003
Advance Access Publication 8 February 2024
Research Article

Exposure to air pollution is associated with DNA methylation changes in sperm

Rose Schrott[1,2,‡], Jason I. Feinberg[1,2,‡], Craig J. Newschaffer[3], Irva Hertz-Picciotto[4], Lisa A. Croen[5], M. Daniele Fallin[6], Heather E. Volk[1,2,§,*], Christine Ladd-Acosta[1,7,§,*] and Andrew P. Feinberg[2,8,§,*]

[1]Wendy Klag Center for Autism and Developmental Disabilities, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21205, USA, [2]Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21205, USA, [3]Department of Biobehavioral Health, College of Health and Human Development, Pennsylvania State University, State College, PA 16802, USA, [4]Department of Public Health Sciences, MIND (Medical Investigations of Neurodevelopmental Disorders) Institute, University of California, Davis, CA 95616, USA, [5]Division of Research, Kaiser Permanente Northern California, Oakland, CA 94612, USA, [6]Rollins School of Public Health, Emory University, Atlanta, GA 30322, USA, [7]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21205, USA, [8]Center for Epigenetics, Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA
[‡]These authors contributed equally.
[§]These authors jointly supervised this work.
*Correspondence address. Johns Hopkins University, 624 N. Broadway, Hampton House 833, Baltimore, MD 21205, USA. E-mail: hvolk1@jhu.edu
Johns Hopkins University, 615 N. Wolfe Street, Room W6509, Baltimore, MD 21205, USA. E-mail: claddac1@jhu.edu
Johns Hopkins University, 855 N. Wolfe Street, Rangos 570, Baltimore, MD 21205, USA. E-mail: afeinberg@jhu.edu

Abstract

Exposure to air pollutants has been associated with adverse health outcomes in adults and children who were prenatally exposed. In addition to reducing exposure to air pollutants, it is important to identify their biologic targets in order to mitigate the health consequences of exposure. One molecular change associated with prenatal exposure to air pollutants is DNA methylation (DNAm), which has been associated with changes in placenta and cord blood tissues at birth. However, little is known about how air pollution exposure impacts the sperm epigenome, which could provide important insights into the mechanism of transmission to offspring. In the present study, we explored whether exposure to particulate matter less than 2.5 microns in diameter, particulate matter less than 10 microns in diameter, nitrogen dioxide (NO_2), or ozone (O_3) was associated with DNAm in sperm contributed by participants in the Early Autism Risk Longitudinal Investigation prospective pregnancy cohort. Air pollution exposure measurements were calculated as the average exposure for each pollutant measured within 4 weeks prior to the date of sample collection. Using array-based genome-scale methylation analyses, we identified 80, 96, 35, and 67 differentially methylated regions (DMRs) significantly associated with particulate matter less than 2.5 microns in diameter, particulate matter less than 10 microns in diameter, NO_2, and O_3, respectively. While no DMRs were associated with exposure to all four pollutants, we found that genes overlapping exposure-related DMRs had a shared enrichment for gene ontology biological processes related to neurodevelopment. Together, these data provide compelling support for the hypothesis that paternal exposure to air pollution impacts DNAm in sperm, particularly in regions implicated in neurodevelopment.

Key words: air pollution; sperm; DNA methylation; epigenetics; genome-scale

Introduction

Prenatal exposure to air pollution is an environmental risk factor for adverse health outcomes in offspring including neurodevelopmental disorders such as autism, cardiometabolic conditions, and immune dysregulation [1–4]. For example, increased prenatal exposure to particulate matter less than 2.5 microns in diameter ($PM_{2.5}$) is associated with an increased risk of autism in children, with stronger associations found in boys compared to girls [5]. Prenatal exposure to outdoor air pollutants was associated with delayed physical growth in the early childhood years [3]. Exposure to higher levels of traffic-related air pollutants including nitrogen dioxide (NO_2), $PM_{2.5}$, and particulate matter less than 10 microns in diameter (PM_{10}) during pregnancy and early infancy was separately associated with increased autism risk [6], as well as altered nasal mucosal and immune blood profiles during infancy, and increased risk of asthma and allergic rhinitis at age 6 years [4]. More broadly, prenatal exposure to air pollution has been associated with delays in cognitive development, deficits in attention and memory, and brain volume changes in offspring [7, 8]. In mice, prenatal exposure to levels of concentrated ambient ultrafine particles, mimicking high traffic areas of the USA, was associated with neurodevelopmental toxicity, specifically in male offspring [9].

Received 9 November 2023; revised 10 January 2024; accepted 2 February 2024

Air pollution exposure in men has been associated with DNA fragmentation in sperm [10], reduced semen quality [11], decreased sperm motility [12], and aberrant sperm motion [12]. A recent systematic meta-analysis reported that outdoor air pollution was significantly associated with a suite of impaired semen parameters including decreased semen volume, low sperm concentration, impaired progressive and total motility, and abnormal sperm morphology rates [13]. These findings are consistent for both gaseous and particulate matter pollutants. In mice, simulated wildfire smoke was found to significantly alter sperm DNA methylation (DNAm) patterns [14].

Despite studies demonstrating the deleterious effects of air pollution exposure on sperm, little is known about how air pollution exposure impacts the human sperm epigenome and whether this might be an important paternal environmental risk factor for health outcomes in their offspring. This is relevant given that the sperm DNA methylome is impacted by numerous environmental and lifestyle exposures. For example, there are known age-related changes to the sperm methylome [15], and exposures such as obesity [16, 17], high-fat diet [18], tobacco [19], cannabis [20, 21], and endocrine disrupting compounds [22] have all been associated with sperm DNAm differences. Furthermore, there is evidence to support that sperm DNAm is important for and associated with developmental outcomes in children. Previously in the Early Autism Risk Longitudinal Investigation (EARLI) cohort, we demonstrated that sperm DNAm was associated with offspring quantitative autistic traits when children were 12 [23] and 36 months old [24]. However, to our knowledge, the impact of air pollution on the human sperm DNAm has not been rigorously investigated. Expanding our understanding of whether air pollution influences the human sperm DNA methylome will help us understand whether these gametic epigenetic modifications are a plausible mechanism through which exposure to air pollution might influence child health outcomes.

The current study was conducted in the EARLI cohort. EARLI is a prospective pregnancy cohort focused on the younger sibling of a child already diagnosed with autism. Given the high rate of sibling recurrence for autism, there is an increased likelihood that the younger sibling will receive an autism diagnosis [25]. Previously in EARLI, we demonstrated that prenatal air pollution exposure was associated with epigenetic changes in cord blood and placenta tissues [26]; however, the impact of this exposure on paternal epigenetic changes remains largely unknown. To address this gap, the primary objective of this study was to determine whether exposure to the traffic-related air pollutants $PM_{2.5}$, PM_{10}, NO_2, or ozone (O_3) is associated with DNAm in sperm in EARLI.

Results

Analytic sample description

Bivariate associations between study population characteristics and the four air pollutant measures are presented in Table 1. Fathers were predominantly White (76%), and their ages ranged from 27.6 to 49.9 years, with an average age of 36.3 years (Table 1). There were no significant differences in the mean exposure level by paternal smoking status, education level, or self-reported race for any of the four air pollutants. There was a significant inverse association between paternal age and PM_{10} exposure ($P = 0.0001$), where younger fathers had higher levels of exposure and NO_2 exposure was higher among fathers living on the east coast ($P < 0.05$; Table 1).

The average [mean (SD)] levels of $PM_{2.5}$, PM_{10}, NO_2, and O_3 exposure were 10.58 $\mu g/m^3$ (4.13), 19.06 $\mu g/m^3$ (5.55), 12.99 ppb (5.11), and 40.43 ppm (12.77), respectively (Table 2). According to the National Ambient Air Quality Standards for these four pollutants, these 4-week averages are below the Environmental Protection Agenency (EPA) standards for these pollutants. $PM_{2.5}$ and PM_{10} were moderately but significantly correlated with one another ($r = 0.43$), as were PM_{10} and O_3 ($r = 0.57$) and NO_2 and O_3 ($r = -0.61$) (Table 2).

$PM_{2.5}$-associated differentially methylated regions in sperm

We identified 3996 differentially methylated regions (DMRs) in sperm associated with $PM_{2.5}$ levels, of which 80 reached genome-scale statistical significance (family-wise error rate (FWER) $P < 0.05$, Fig. 1, left). When examining the directionality of the relationship between $PM_{2.5}$ exposure and DNAm, we observed that for 71 (89%) of the 80 significant DMRs, higher levels of $PM_{2.5}$ exposure were associated with less DNAm, i.e. hypomethylation, in sperm. The genes annotated to the three most significant DMRs were *O*-6-methylguanine-DNA methyltransferase (*MGMT*), a gene involved in DNA repair mechanisms [27]; SET and MYND domain-containing 3 (*SMYD3*), which is a histone methyltransferase involved in chromatin organization and transcriptional regulation [28]; and POU Class 6 Homeobox 2 (*POU6F2*), a transcriptional regulator involved in cellular differentiation [29] (Fig. 1, right). For both *MGMT* and *POU6F2*, higher levels of $PM_{2.5}$ exposure were associated with hypomethylated DNA, while the inverse relationship was observed for *SMYD3*. The average magnitudes of methylation difference across the highest and lowest quartiles of $PM_{2.5}$ exposure were 7.5%, 14.3%, and 14.6% for *MGMT, SMYD3*, and *POU6F2*, respectively. The top 10 DMRs associated with $PM_{2.5}$ are shown in Table 3a (complete DMR list given in Supplementary Table S2).

PM_{10}-associated DMRs in sperm

We next examined the relationship between PM_{10} exposure and sperm DNAm. There were 2439 DMRs in sperm associated with PM_{10} exposure, of which 96 met our statistical significance threshold (FWER $P < 0.05$, Fig. 2, left). When assessing proportions of hyper- and hypomethylated DMRs in response to PM_{10} exposure, we observed that for 71% of DMRs, higher levels of PM_{10} exposure were associated with lower levels of methylation in sperm. This relationship is demonstrated by the patterns observed for the three most significant DMRs, as shown in Fig. 2 (right-hand side). For all three genes—MAGE Family Member E2 (*MAGEE2*), Chymase 1 (*CMA1*), and RNA-Binding Fox-1 Homolog 1 (*RBFOX1/A2BP1*)—an inverse relationship was observed, where higher levels of PM_{10} exposure were associated with hypomethylated DNA in sperm. The magnitudes of methylation difference in sperm from men exposed to the highest and lowest quartiles of PM_{10} exposure were also notable and showed a dose–response relationship. For example, when comparing men exposed to the highest quartile of PM_{10} compared to the lowest quartile of PM_{10}, the largest losses in methylation were observed at DMRs overlapping the genes Catenin Alpha 2 (*CTNNA2*), Neurexin 1 (*NRXN1*), and *A2BP1*, where sperm DNA was hypomethylated by 29.5%, 19.4%, and 19.2%, respectively. Looking across all four quartiles of PM_{10} exposure for *A2BP1*, we observed an example of a dose–response relationship, where Q1, Q2, Q3, and Q4 methylation levels were 68.24%, 58.74%, 54.32%, and 50.10%, respectively. Interestingly, all three of these genes play important roles in neurodevelopment and synaptic function [30–32]. The top 10 DMRs associated with PM_{10} are shown in Table 3b (complete DMR list given in Supplementary Table S3).

Table 1: Bivariate associations between study sample characteristics and air pollution

		PM2.5		PM10		NO_2		O_3	
	n (%)	Mean (SD)	*P*	Mean (SD)	*P*	Mean (SD)	*P*	Mean (SD)	*P*
Age [Mean (SD)]	36.34 (5.56)		0.298916097		0.000107154		0.450737823		0.445602849
Race			0.203636859		0.195209062		0.711225136		0.266359125
White	29 (76.3)	10.33 (3.49)		18.31 (4.83)		13.14 (5.54)		40.43 (12.85)	
Black	4 (10.5)	10.68 (4.69)		21.78 (3.01)		12.27 (4.68)		49.85 (14.44)	
Asian	2 (5.3)	7.48 (2.81)		16.58 (1.3)		9.5 (0.67)		35.07 (4.78)	
Other	3 (7.9)	14.97 (8.33)		24.44 (12.45)		14.88 (0.74)		31.47 (7.32)	
Study site			0.09531417		0.063383362		0.045152783		0.530528522
Drexel	14 (36.8)	11.17 (4.05)		20.91 (4.92)		13.79 (5.2)		44.52 (13.75)	
Johns Hopkins	11 (28.9)	11.89 (2.05)		16.38 (3.35)		15.33 (4.64)		38.21 (15.52)	
Kaiser	6 (15.8)	6.87 (1.85)		16.52 (4.33)		8.62 (4.6)		37.37 (7.16)	
UC Davis	7 (18.4)	10.54 (6.5)		21.77 (8.18)		11.45 (3.9)		38.4 (9.43)	
Coast			0.85		0.11		0.009		0.31
East	25 (65.8)	11.49 (3.28)		18.92 (4.80)		14.47 (4.92)		41.74 (14.60)	
West	13 (34.2)	8.85 (5.12)		19.43 (6.98)		10.15 (4.31)		37.92 (8.13)	
Paternal smoking (ever)			0.442107804		0.730216088		0.615353619		0.25279209
Yes	18 (47.4)	10.61 (3.37)		18.4 (5.26)		13.03 (5.11)		42.7 (13.66)	
No	15 (39.5)	9.6 (4)		17.85 (3.81)		13.93 (5.03)		37.57 (11.6)	
Missing	5 (13.2)	13.44 (6.32)		25.1 (7.93)		10.02 (5.25)		40.88 (13.76)	
Paternal education			0.143268847		0.196401961		0.496795552		0.879850982
Less than High School (HS)	—	—		—		—		—	
HS/General Education	7 (18.4)	12.75 (6.66)		22.93 (7.53)		11.29 (4.17)		43.43 (15.73)	
Development diploma									
Some college	8 (21.1)	9.67 (3.05)		19.21 (6.3)		11.94 (4.51)		41.31 (12.12)	
Bachelor's degree	10 (26.3)	8.56 (1.93)		17.3 (2.24)		12.9 (4.66)		40.08 (10.94)	
Graduate/professional degree	13 (34.2)	11.53 (3.82)		18.25 (5.25)		14.63 (6.18)		38.55 (13.92)	
PM2.5.ppb.4weekAvg [mean (SD)]	10.58. (4.13)		0		0.006973339		0.107692615		0.283176209
PM10.ppb.4weekAvg [mean (SD)]	19.06 (5.55)		0.006973339		0		0.105911596		0.000235082
NO2.ppb.4weekAvg [mean (SD)]	12.99 (5.11)		0.107692615		0.105911596		0		6.99×10^{-5}
O3.ppb.4weekAvg [mean (SD)]	40.43 (12.77)		0.283176209		0.000235082		6.99×10^{-5}		0

For continuous variables, Spearman correlation tests were performed. For dichotomous variables, the Mann–Whitney/Wilcoxon rank-sum test was performed. For nominal variables, the Kruskal–Wallis test was performed. Numbers reported in the first column are the number and corresponding percent of the analytic population unless otherwise specified.

Table 2: Between-pollutant correlations. Spearman correlations across pollutants examined

	Mean (SD)	$PM_{2.5}$	PM_{10}	NO_2	O_3
$PM_{2.5}$	10.58. (4.13)	1 (0.00)	0.4336361 (0.007)	0.2651275 (0.108)	0.1782471 (0.283)
PM_{10}	19.06 (5.55)	0.4336361 (0.007)	1 (0.00)	–0.2664405 (0.106)	0.5708502 (0.00024)
NO_2	12.99 (5.11)	0.2651275 (0.108)	–0.2664405 (0.106)	1 (0.00)	–0.6091476 (6.99E-05)
O_3	40.43 (12.77)	0.1782471 (0.283)	0.5708502 (0.00024)	–0.6091476 (6.99E-05)	1 (0.00)

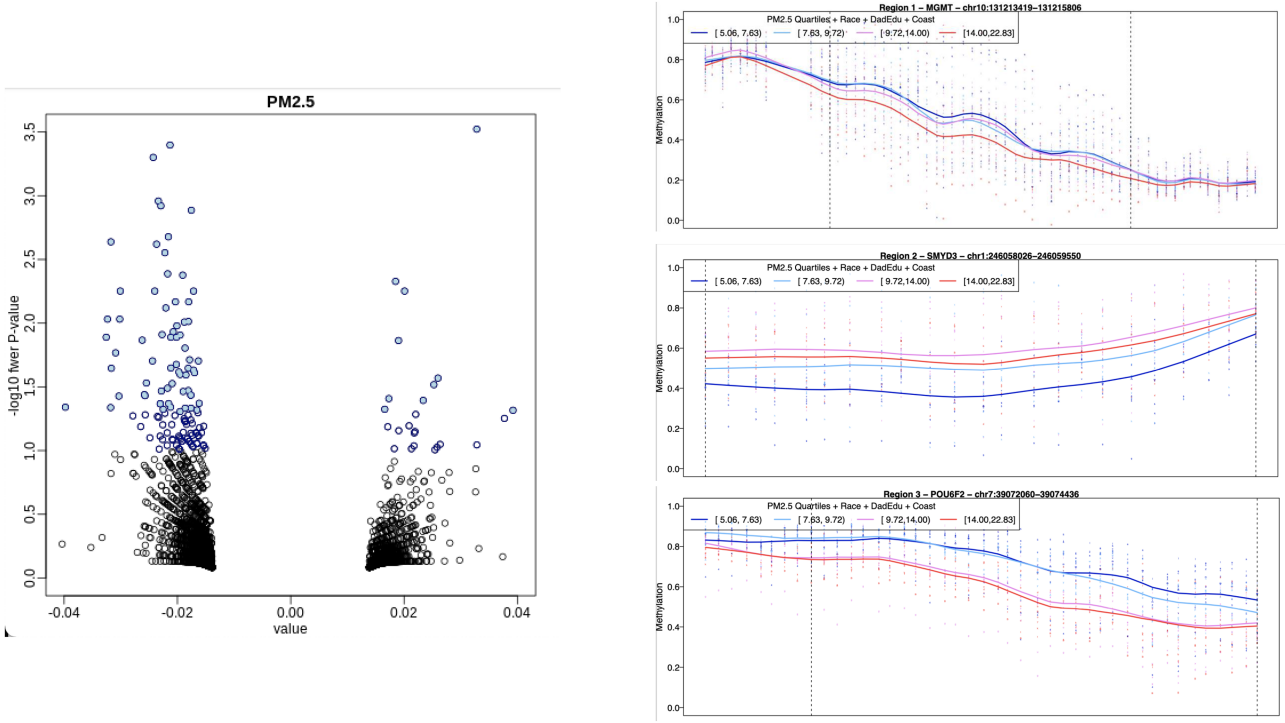

Figure 1: (Left) Volcano plots for 3996 PM2.5 DMRs Y-axis shows the –log10(FWER P) for each DMR returned by the bump hunter algorithm after 10 000 bootstrap permutations. X-axis is the CHARM DMR value, which corresponds to the smoothed effect estimate per DMR returned by bump hunter. Filled-in blue circles have FWER $P < 0.05$, open blue circles have FWER $P < 0.1$, and black circles have no nominal significance. There is no stratification by quartile. (Right) Methylation plots for the top three statistical DMRs ($P < 1.0 \times 10^{-4}$) identified using CHARM and PM2.5 exposure levels. The name of the gene to which the DMR is annotated is at the top of each panel. (top) *MGMT*, (middle) *SMYD3*, and (bottom) *POU6F2*. Panels show individual methylation levels at each probe by genomic position. Dotted vertical black lines represent the boundaries of the DMR, and colored lines represent the average methylation curve for samples grouped by quartiles of PM2.5 exposure—the exposure quartiles within each quartile are shown in the legend. The vertical colored dots represent the individual methylation levels for each individual at each genomic position. The direction of the colored lines demonstrates the direction of methylation change across the genomic positions captured by the DMR

NO_2-associated DMRs in sperm

We identified 1649 sperm DMRs associated with NO_2 exposure, with 35 reaching genome-scale statistical significance (FWER $P < 0.05$, Fig. 3, left), 63% of which were hypomethylated following exposure to the highest levels of NO_2. The three most significant sperm DMRs associated with NO_2 exposure are displayed in right panel of Fig. 3, and the top 10 DMRs associated with NO_2 are listed in Table 3c (complete DMR list given in Supplementary Table S4). For the DMRs located in *C7orf4* and C1D nuclear receptor corepressor (*C1D*), higher levels of NO_2 exposure were associated with lower levels of DNAm in sperm. The opposite relationship was observed for the DMR located in the Family with Sequence Similarity 13 Member A (*FAM13A*) gene. The largest magnitude methylation differences reached 31.1% hypomethylation in the Myosin IIIB (*MYO1B*) gene in sperm from individuals exposed to the highest levels of NO_2 compared to the lowest and 36.6% hypermethylation in the LHFPL Tetraspan Subfamily Member 6 (*LHFP*) gene from individuals exposed to the highest levels of NO_2 relative to those exposed to the lowest. *FAM13A* is reported to be involved in airway epithelium remodeling, with suspected roles in chronic obstructive pulmonary disease [33]; *MYOB1* has functions important for cytoskeletal motor activity [34]; and *LHFP* is a member of the LHFP superfamily of tetraspan transmembrane protein encoding genes, with a suspected role in gastric cancer [35].

O_3-associated DMRs in sperm

We identified 1922 sperm DMRs associated with O_3 exposure levels around the time of conception, of which 67 reached a statistical significance threshold (FWER $P < 0.05$, Fig. 4, left). More than half of the DMRs (55%) showed decreases in methylation associated with higher levels of O_3 exposure. The three most significant DMRs are shown in the right panel of Fig. 4, and the top 10 DMRs associated with O_3 are shown in Table 3d (complete DMR list given in Supplementary Table S5). For all three of these genes—Catenin Delta 1 (*CTNND1*), *LOC646982*, and MLLT3 Super Elongation Complex Subunit (*MLLT3*)—higher levels of O_3 exposure were associated with increased methylation in sperm. For many of the DMRs, large

Table 3: The top 10 significant DMRs in paternal sperm associated with (a) $PM_{2.5}$, (b) PM_{10}, (c) NO_2, and (d) O_3

Genomic location	FWER P-value	Symbol	Genic location	Distance	DeltaM	Function
(a) Top 10 $PM_{2.5}$-associated DMRs						
chr10:131213419-131215806	0.0001	*MGMT*	Upstream	49647	−0.075	Involved in DNA repair of methylated nucleobases
chr1:246058026-246059550	0.0003	*SMYD3*	Inside intron	521160	0.143	Histone methyltransferase (H3K4me2/3). Functions as a transcriptional activator
chr7:39072060-39074436	0.0004	*POU6F2*	Inside intron	54452	−0.146	Probable transcription factor likely to be involved in early steps in the differentiation of amacrine and ganglion cells
chr21:31399849-31400846	0.0004	*GRIK1*	Upstream	87568	−0.208	Ionotropic glutamate receptor (excitatory neurotransmitter receptor)
chr12:130669848-130671949	0.0005	*FZD10*	Downstream	22817	−0.115	Member of the frizzled gene family. Wnt protein receptor
chr1:234905850-234907810	0.0011	*IRF2BP2*	Upstream	160580	−0.159	Encodes an interferon regulatory factor-2 (IRF2) binding protein that interacts with the C-terminal transcriptional repression domain of IRF2
chrX:68382003-68383827	0.0012	*PJA1*	Overlaps exon upstream	1444	−0.154	Has E2-dependent E3 ubiquitin-protein ligase activity
chr12:117296502-117298878	0.0013	*HRK*	Downstream	20353	−0.069	Member of the BCL-2 protein family. Members of this family are involved in activating apoptosis
chr2:3011215-3013080	0.0021	*TSSC1/EIPR1*	Downstream	368572	−0.071	Component of endosomal retrieval machinery involved in the trans-Golgi network for protein recycling. Located in the imprinted gene domain 11p15.5
chr2:68132210-68133404	0.0023	*C1D*	Downstream	156754	−0.128	DNA-binding and apoptosis-inducing protein, localized in the nucleus. Rac3-interacting protein that acts as a corepressor for the thyroid hormone receptor
(b) Top 10 PM_{10}-associated DMRs						
chrX:74961420-74964310	<0.0001	*MAGEE2*	Downstream	40760	−0.117	Tumor antigen
chr14:24974565-24976858	<0.0001	*CMA1*	Overlaps 3'	612	−0.092	Protease of mast cells involved in vasoactive peptide generation, extracellular matrix degradation, and regulation of gland secretion
chr16:7142674-7144522	<0.0001	*A2BP1*	Inside intron	318865	−0.182	RNA-binding protein that regulates alternative splicing events
chr10:3057233-3058630	<0.0001	*PFKP*	Upstream	51121	0.03	Involved in glycolysis
chr19:30745990-30748340	0.0001	*ZNF536*	Upstream	114987	0.112	May be involved in transcriptional regulating; metal binding
chr5:31681937-31683165	0.0005	*PDZD2*	Upstream	115865	0.163	Involved in cell adhesion and prostate tumorigenesis
chr8:136878647-136880446	0.0005	*KHDRBS3*	Downstream	408932	−0.122	RNA-binding protein that plays a role in regulating alternative splicing and influences mRNA splice site selection and exon exclusion. Can regulate alternative splicing of *NRXN1-3* that functions to target neurexins to postsynaptic partners in neuroligins and leucine-rich transmembrane protein family members
chr13:101536306-101537485	0.0005	*NALCN*	Downstream	531327	0.155	Voltage-independent, nonselective cation channel which belongs to a family of voltage-gated sodium and calcium channels that regulate the resting membrane potential and excitability of neurons; expressed throughout the nervous system and conducts a persistent sodium leak current that contributes to tonic neuronal excitability
chr2:120011056-120012516	0.0007	*STEAP3*	Covers exon(s)	29673	0.06	Oxidoreductase and iron transporter; mediates responses to p53, and promotes apoptosis
chr10:13868030-13869444	0.0007	*FRMD4A*	Inside intron	503421	−0.118	Regulates the remodeling of adherens junctions and linear actin cable formation during epithelial cell polarization. Polymorphisms are associated with Alz
(c) Top 10 NO_2-associated DMRs						
chr7:155988735-155991583	0.0002	*C7orf4*	Upstream	341601	−0.177	RNA gene is affiliated with the lncRNA class
chr4:89679783-89681417	0.0015	*FAM13A*	Covers exon(s)	62984	0.128	Predicted to enable GTPase activator activity

(continued)

Table 3: (Continued)

Genomic location	FWER P-value	Symbol	Genic location	Distance	DeltaM	Function
chr2:68132210-68133404	0.0017	*C1D*	Downstream	156754	−0.137	DNA-binding and apoptosis-inducing protein, localized in the nucleus. Rac3-interacting protein that acts as a corepressor for the thyroid hormone receptor
chr21:31399849-31400846	0.0045	*GRIK1*	Upstream	87568	−0.229	Ionotropic glutamate receptor (excitatory neurotransmitter receptor)
chr20:34458591-34459341	0.0045	*PHF20*	Covers exon(s)	98669	−0.095	Methyllysine-binding protein, chromatin regulator component of the MOF histone acetyltransferase protein complex
chr18:73608225-73609747	0.0068	*ZNF516*	Downstream	565307	−0.115	Transcriptional regulator that promotes brown adipose tissue differentiation
chr8:4697089-4698489	0.0103	*CSMD1*	Inside intron	153838	0.167	Predicted to act upstream of many processes including learning/memory and reproductive structure development
chr11:57639658-57641171	0.0109	*CTNND1*	Downstream	110425	−0.148	Armadillo protein family member involved in cell–cell adhesion
chr9:129642026-129643484	0.0109	*ZBTB34*	Inside exon	19083	−0.146	Predicted to enable DNA-binding transcription repressor activity
chr12:77965737-77966649	0.0112	*NAV3*	Upstream	258419	−0.294	Part of the neuron navigator family of proteins that is predominantly expressed in the nervous system
(d) Top 10 O_3-associated DMRs						
chr11:57639658-57641253	<0.0001	*CTNND1*	Downstream	110425	0.308	Armadillo protein family member involved in cell–cell adhesion
chr13:40976148-40978047	<0.0001	*LOC646982*	Inside intron	77095	0.218	Long noncoding RNA
chr9:20267374-20268174	0.0004	*MLLT3*	Downstream	354339	0.377	chromatin reader component of the super elongation complex that increases the catalytic rate of RNA pol II txn
chr7:155989653-155991513	0.0015	*C7orf4; LINC00244*	Upstream	341671	0.077	RNA gene is affiliated with the lncRNA class
chr8:3722198-3723528	0.0015	*CSMD1*	Inside intron	1128799	−0.201	Predicted to act upstream of many processes including learning/memory and reproductive structure development
chr1:15066168-15067918	0.0041	*KIAA1026*	Inside intron	140956	−0.124	Gene encodes a protein that plays a role in desmosome assembly, cell adhesion, cytoskeletal organization, and epidermal differentiation
chr1:215843109-215844569	0.0042	*USH2A*	Overlaps exon downstream	752168	−0.108	Involved in hearing and vision as member of the USH2 complex. Protein is found in the basement membrane and may be important in development and homeostasis of the inner ear and retina.
chr4:186693264-186695194	0.0047	*SORBS2*	Inside intron	1871	−0.119	Adapter protein that plays a role in the assembling of signaling complexes, being a link between ABL kinases and actin cytoskeleton.
chr12:77965737-77966707	0.0050	*NAV3*	Upstream	258361	0.29	This gene belongs to the neuron navigator family and is expressed predominantly in the nervous system. May regulate *interleukin-2* production by T-cells. May be involved in neuron regeneration.
chr16:78376735-78378365	0.0055	*WWOX*	Inside intron	243185	−0.112	Oxidoreductase involved in neuronal signaling and DNA damage repair

The boundaries of the DMR are shown in the genomic location column; the FWER P-value is displayed alongside the gene symbol. The genic location characterized where within the gene body the DMR is located. deltaM is the difference in sperm methylation from the highest and lowest exposure quartiles. Gene functions were taken from the human protein atlas (proteinatlas.org) as well as gene cards (genecards.org).

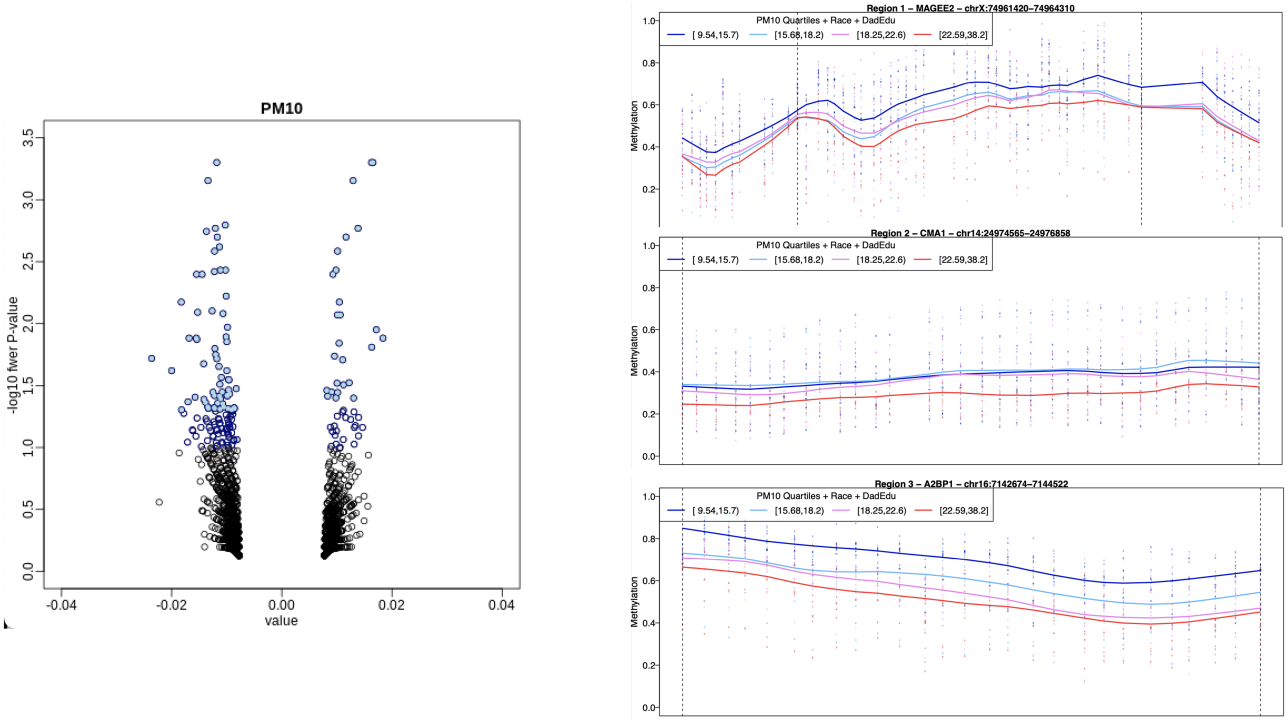

Figure 2: (Left) Volcano plots for 2439 PM10 DMRs Y-axis show the –log10(FWER P) for each DMR returned by the bump hunter algorithm after 10 000 bootstrap permutations. X-axis is the CHARM DMR value which corresponds to the smoothed effect estimate per DMR returned by bump hunter. Filled-in blue circles have FWER $P < 0.05$, open blue circles have FWER $P < 0.1$, and black circles have no nominal significance. There is no stratification by quartile. (Right) Methylation plots for the top three statistical DMRs ($P < 1.0 \times 10^{-4}$) identified using CHARM and PM10 exposure levels. The name of the gene to which the DMR is annotated is at the top of each panel. (Top) *MAGEE2*, (middle) *CMA1*, and (bottom) *A2BP1*. Panels show individual methylation levels at each probe by genomic position. Dotted vertical black lines represent the boundaries of the DMR, and colored lines represent the average methylation curve for samples grouped by quartiles of PM10 exposure—the exposure quartiles within each quartile are shown in the legend. The vertical colored dots represent the individual methylation levels for each individual at each genomic position. The direction of the colored lines demonstrates the direction of methylation change across the genomic positions captured by the DMR

magnitudes of change in sperm methylation levels were observed in men exposed to the highest levels of O_3 compared to those with the lowest exposure. For example, in the gene Diacylglycerol Kinase Gamma (*DGKG*), sperm from men exposed to the highest levels of O_3 was 35.5% hypomethylated relative to sperm from men exposed to the lowest levels of O_3. In the gene Mannose Binding Lectin 2 (*MBL2*), sperm DNAm from men exposed to the highest O_3 levels was hypermethylated by 39.6% relative to those exposed to the lowest O_3 level. Interestingly, there were four different intronic regions of the gene CUB and Sushi Multiple Domains 1 (*CSMD1*) that were significantly associated with O_3 exposure, of which the majority were hypomethylated with higher exposures. This gene is expressed in multiple brain regions, and it has predicted functions in neurobehavioral processes such as learning and memory [36].

Cross-platform validation of our air pollutant-associated DMRs

To validate the DNAm changes, we observed using measurements from the Comprehensive High-throughput Arrays for Relative Methylation (CHARM) array and we used methylation data from an independent platform: the Illumina 450 K array, a highly reproducible methylation measurement method. Supplementary Table S6a–d shows the degree of association between 450 K variables of each of the estimated surrogate variables (SVs) for all four pollutants. We extracted 450 K methylation probes located within 500 base pairs (bp) of the pollutant-associated CHARM DMR boundaries. This was possible for 39 (48.75%) of the 80 DMRs identified for $PM_{2.5}$, 30 (31.25%) of the 96 PM_{10}-associated DMRs, 13 (37%) of 35 NO_2, and 19 (28.36%) of 67 O_3. The direction of association between the pollutant and DNAm was consistent across the two platforms for 36 of the 39 $PM_{2.5}$-associated regions (rho = 0.34, Supplementary Fig. S1a), 28 of 30 PM_{10}-associated regions (rho = 0.75, Supplementary Fig. S1b), 5 of 13 NO_2-associated regions (rho = –0.06, Supplementary Fig. S1c), and 17 of 19 O_3-associated regions (rho = 0.76, Supplementary Fig. S1d).

Cross-pollutant analyses

Comparisons across all pollutants revealed no DMRs that were associated with all four pollutants, although there were DMRs that were in common across two or three of the pollutants (Fig. 5). Despite the pollutant estimates themselves not being correlated with each other, there were 14 DMRs in common between $PM_{2.5}$ and O_3, including multiple genes involved in learning memory and neurodevelopment such as *CSMD1*, *CTNNA2*, and Glutamate Ionotropic Receptor Kainate Type Subunit 1 (*GRIK1*). Between NO_2 and O_3, two pollutants that were correlated with each other, there were 13 DMRs in common, including three that were similarly associated with PM_{10}: *MLLT3*, *CTNND1*, and Neuron Navigator 3 (*NAV3*), which are involved in early developmental patterning and nervous system development. There were 12 DMRs in common between PM_{10} and O_3, the only two pollutants from different classes (particulate matter vs. gaseous) that were correlated with one another. $PM_{2.5}$ and PM_{10} were similarly correlated with one another and shared eight DMRs that were associated with both exposures.

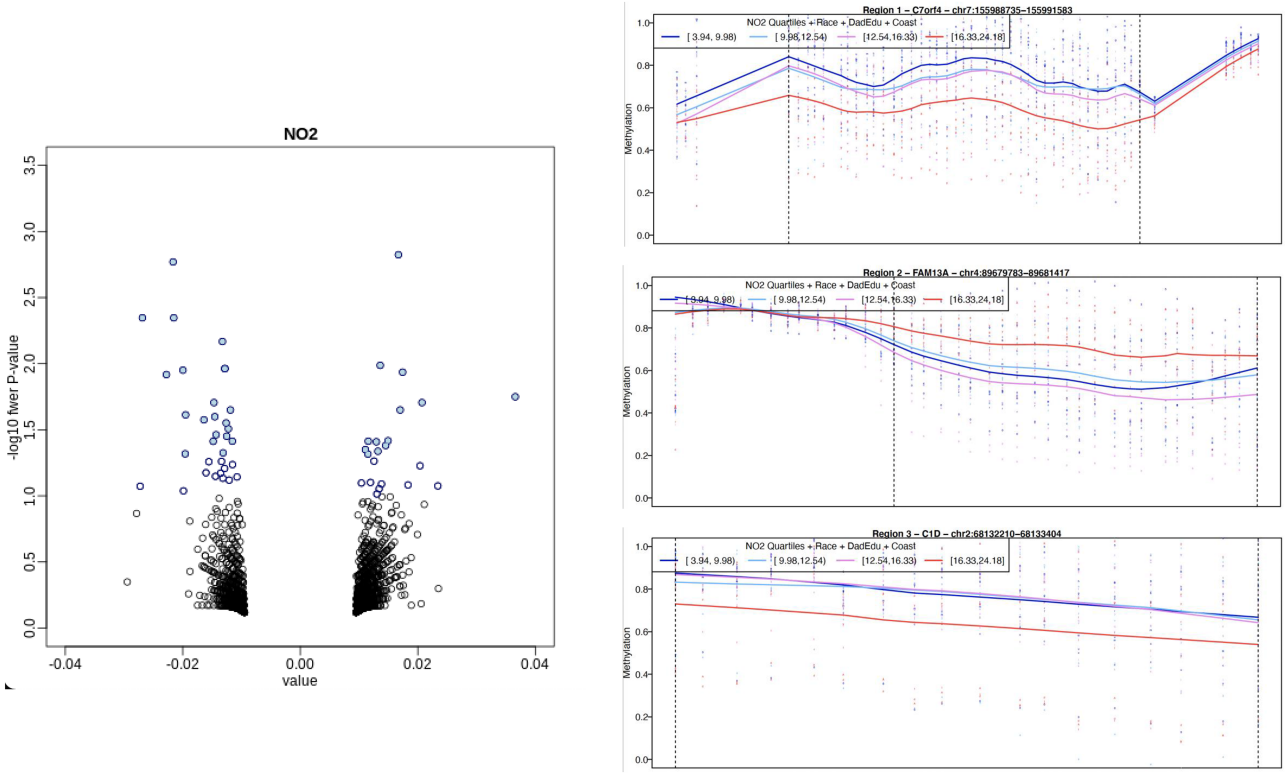

Figure 3: (Left) Volcano plots for 1649 NO_2 DMRs Y-axis shows the –log10(FWER P) for each DMR returned by the bump hunter algorithm after 10 000 bootstrap permutations. X-axis is the CHARM DMR value which corresponds to the smoothed effect estimate per DMR returned by bump hunter. Filled-in blue circles have FWER $P < 0.05$, open blue circles have FWER $P < 0.1$, and black circles have no nominal significance. There is no stratification by quartile. (Right) Methylation plots for the top three statistical DMRs ($P < 1.0 \times 10^{-4}$) identified using CHARM and NO_2 exposure levels. The name of the gene to which the DMR is annotated is at the top of each panel. (Top) *C7orf4*, (middle) *FAM13A*, and (bottom) *C1D*. Panels show individual methylation levels at each probe by genomic position. Dotted vertical black lines represent the boundaries of the DMR, and colored lines represent the average methylation curve for samples grouped by quartiles of NO_2 exposure—the exposure quartiles within each quartile are shown in the legend. The vertical colored dots represent the individual methylation levels for each individual at each genomic position. The direction of the colored lines demonstrates the direction of methylation change across the genomic positions captured by the DMR

When comparing the proportion of DMRs that were hyper- or hypomethylated in response to increasing levels of air pollutant, we observed differences in the patterns observed between particulate matter and gaseous pollutants. While the majority of DMRs for all four pollutants showed hypomethylation association with increasing levels of air pollution, the proportion of hypomethylated DMRs for NO_2 (63%) and O_3 (55%) was substantially less than that for $PM_{2.5}$ (89%) and PM_{10} (71%).

Gene ontology analyses

We performed gene enrichment analyses to identify biological processes related to genes overlapping DMRs (FWER $P < 0.1$, located within 10 kb of the gene) that were associated with each exposure. We identified 59 gene ontology (GO) Biological Process terms associated with genes annotated to DMRs that were associated with $PM_{2.5}$ exposure (Supplementary Table S7a, Fig. S2). Many of the GO terms identified were relevant to neurodevelopment such as "neurogenesis" and "nervous system development." Ontology pathway analyses revealed 32 GO Biological Process terms associated with genes overlapping PM_{10}-associated DMRs (FWER $P < 0.1$, located within 10 kb of the gene). Similar to terms enriched among the $PM_{2.5}$-associated genes, many of these terms were associated with neurogenesis and neurodevelopment, with additional enrichment for terms associated with immune response and activation (Supplementary Table S7b, Fig. S2). There were just two GO Biological Process terms associated with genes annotated to NO_2-associated DMRs (FWER $P < 0.1$, located within 10 kb of the gene, Supplementary Table S7c, Fig. S2): "regulation of transcription by RNA polymerase II" and "central nervous system development." Finally, there were seven GO Biological Process terms enriched for genes overlapping DMRs (FWER $P < 0.1$, within 10 kb of the gene, Supplementary Table S7d, Fig. S2) associated with O_3 exposure, although none of these terms were unique to O_3 exposure.

There were some notable similarities between GO terms across pollutants (Supplementary Fig. S3). For example, there were three GO terms associated with $PM_{2.5}$, PM_{10}, and O_3: "cell adhesion," "biological adhesion," and "cell morphogenesis." Additionally, there were two GO terms associated with $PM_{2.5}$ and O_3, "modulation of chemical synaptic transmission" and "regulation of trans-synaptic signaling," and two separate terms, "muscle cell differentiation" and "negative regulation of response to external stimulus," were associated with both PM_{10} and O_3. We observed the largest overlap in GO terms between $PM_{2.5}$ and PM_{10}, where 14 were in common, including "neurogenesis," "cell–cell adhesion," and "generation of neurons" (Supplementary Fig. S3). The two GO terms annotated to NO_2 were not shared with any other pollutants.

Discussion

The main objective of this study was to identify regions of the genome showing differential methylation in sperm associated

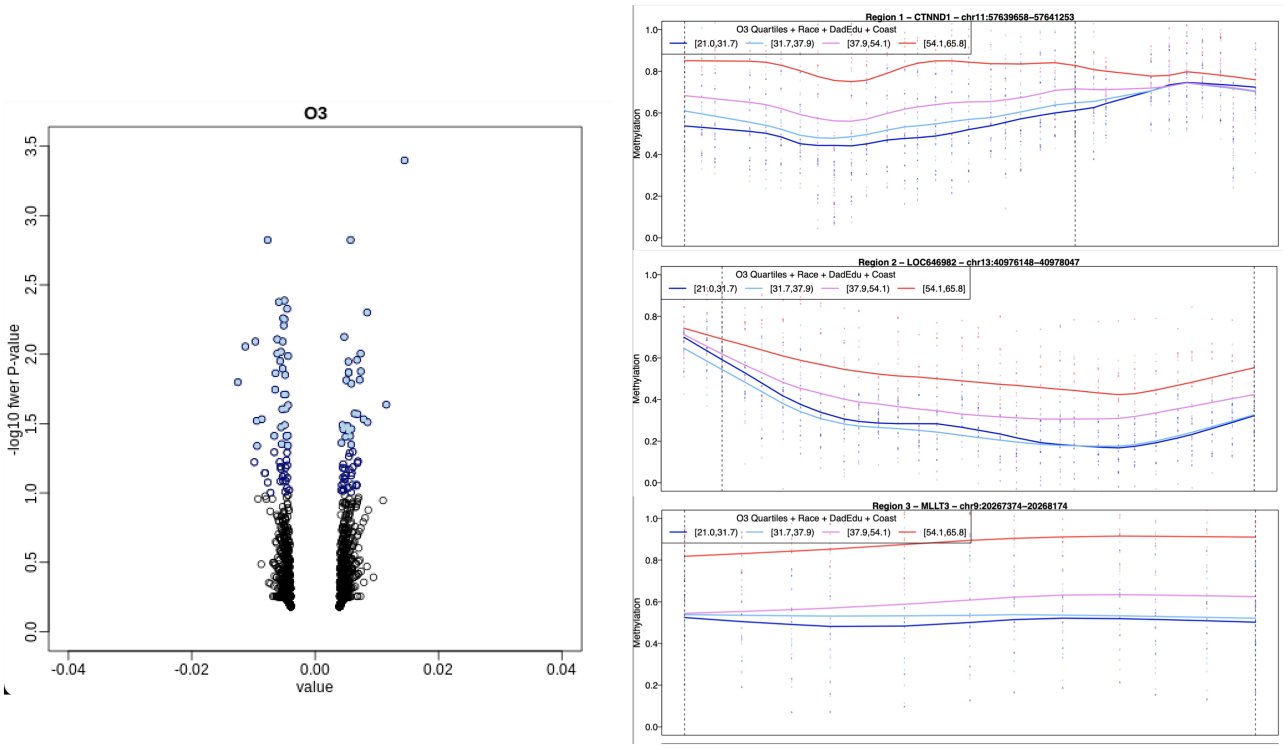

Figure 4: (Left) Volcano plots for 1922 O_3 DMRs Y-axis shows the –log10(FWER P) for each DMR returned by the bump hunter algorithm after 10 000 bootstrap permutations. X-axis is the CHARM DMR value, which corresponds to the smoothed effect estimate per DMR returned by bump hunter. Filled-in blue circles have FWER $P < 0.05$, open blue circles have FWER $P < 0.1$, and black circles have no nominal significance. There is no stratification by quartile. (Right) Methylation plots for the top three statistical DMRs ($P < 1.0 \times 10^{-4}$) identified using CHARM and O_3 exposure levels. The name of the gene to which the DMR is annotated is at the top of each panel. (Top) *CTNND1*, (middle) *LOC646982*, and (bottom) *MLLT3*. Panels show individual methylation levels at each probe by the genomic position. Dotted vertical black lines represent the boundaries of the DMR, and colored lines represent the average methylation curve for samples grouped by quartiles of O_3 exposure—the exposure quartiles within each quartile are shown in the legend. The vertical colored dots represent the individual methylation levels for each individual at each genomic position. The direction of the colored lines across the X-axis demonstrates the direction of methylation change across the genomic positions captured by the DMR

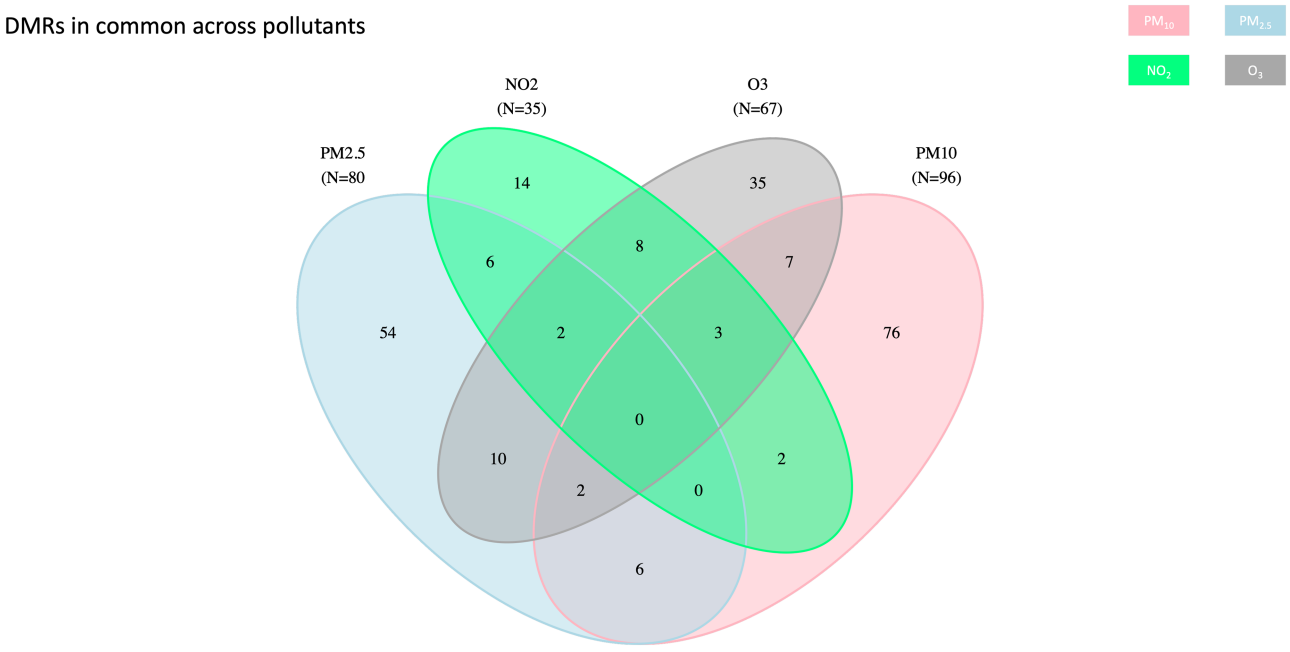

Figure 5: Venn diagram of DMRs associated with one or more pollutant. No DMRs are associated with all four pollutants. $PM_{2.5}$ = blue; PM10 = pink; NO_2 = green; O_3 = gray

with paternal air pollution exposure levels, around the time of conception, in a prospective pregnancy cohort. We identified significant DNAm changes in sperm that were independently associated with each of the four traffic-related pollutants. Air pollution exposures themselves were correlated across pollutants for $PM_{2.5}$ and PM_{10}, PM_{10} and O_3, and NO_2 and O_3, and we identified multiple DMRs in sperm that were associated with two or three of the pollutants. While no individual DMRs overlapped across all four pollutants, we did find that genes involved in neurodevelopmental processes were enriched in all four pollutants. Together, these data provide compelling support for the hypothesis that paternal exposure to air pollution impacts DNAm in sperm, particularly in regions enriched for neurodevelopment. This could provide valuable biologic targets for future investigation into biologic mechanisms through which paternal preconception exposures may influence neurodevelopmental outcomes in their children.

While we are among the first to comprehensively assess the impact of air pollution on the sperm DNA methylome, epidemiologic studies have demonstrated that early life air pollution exposure impacts DNAm at other tissues that are relevant for child development. In a separate analysis conducted in the EARLI cohort, prenatal exposure to NO_2 and O_3 was associated with altered global and locus-specific DNAm in cord blood and placenta tissues [26]. In that study, infants with higher levels of exposure had lower aggregate levels of DNAm [26], which reflects a similar pattern to what we observed in this study where higher levels of air pollution exposure were primarily associated with lower levels of DNAm in sperm. Epidemiologic studies outside of EARLI have also identified DNAm changes in developmental tissues that are associated with prenatal air pollution exposure. In the Healthy Start Study, for example, prenatal $PM_{2.5}$ exposure, particularly during the third trimester, was associated with differential cord blood DNAm at multiple CpG sites, one of which was in the gene *MGMT* [37]. In our study, the DMR in sperm most significantly associated with $PM_{2.5}$ exposure was annotated to the same gene. While literature about the role of *MGMT* in neurodevelopment is sparse, it is well documented that differential DNAm of the *MGMT* promoter is important in predicting clinical response to alkylating drugs used to treat brain tumors such as gliomas [38]. This raises questions about how methylation changes in this gene might be functionally important in neurobiology and in potentially mediating responses to exogenous exposures more broadly.

Genes annotated to DMRs associated with each of the four air pollutants were significantly enriched for biological processes involved in neurodevelopment and neurogenesis. This biological enrichment was despite the fact that no single DMR in sperm was associated with all four exposures, suggesting that there may be underlying biological pathways shared between genes whose methylation patterns are most responsive to air pollution exposure. Some DMRs were identified for more than one air pollutant, and others were uniquely associated with a specific air pollutant. Given the moderate correlation between exposure levels, it is not unexpected that we would find some shared genes, but these results also suggest that there may be pollutant-specific responses in methylation. Indeed, many genes that were associated with two or three of the pollutants, as well as genes that had some of the largest methylation changes associated with the exposures, had functions important for neurodevelopment and synaptogenesis. Furthermore, these findings are in line with other groups that have demonstrated that diet/obesity, cannabis, tobacco smoking, and prenatal exposures previously associated with neurodevelopmental disorders [16, 17, 19–21] have all been associated with altered DNAm at genes implicated in neurodevelopment [17, 19].

One gene of particular interest identified in the present study is *CSMD1*. DMRs from this gene were associated with each of the four pollutants. There were four *CSMD1* DMRs associated with O_3 exposure and one common DMR that was associated with exposure to both NO_2 and O_3. *CSMD1* is expressed at intermediate levels in numerous regions of the brain including the fetal brain, the hippocampus, and the cerebellum [36]. Its functions are strongly implicated in neurogenesis, synaptic scaling, and learning and memory, and genetic variants have been associated with neuropsychiatric conditions such as bipolar disorder and schizophrenia [36]. Furthermore, according to the Simons Foundation for Autism Research Initiative gene database, this gene is a strong autism candidate gene, and genetic variants in *CSMD1* have been identified in individuals with autism spectrum disorder (ASD) from multiplex ASD families [39]. Future studies should investigate the interaction between DNAm and genetic variants in this gene, and whether DNAm changes either in response to or independent of environmental exposures such as air pollution might mediate the effect of genetic variants on neuropsychiatric outcomes.

We previously demonstrated that sperm DNAm in EARLI was associated with autistic traits in fathers and their children at 12 and 36 months of age [23, 24]. The sperm DMRs associated with air pollution exposure in the present study overlapped with those that were associated with autistic traits in children, providing further support that DMRs in sperm that are vulnerable to the environment might influence neurobehavioral traits in children [24]. This lays the groundwork for future studies to investigate mechanisms through which sperm DNAm may mediate the effect of these exposures on neurodevelopmental and ASD-related outcomes.

Methylation measurement technologies can detect DNA sequence variation at or near individual CpGs being interrogated via microarray probes. Types of genetic variants that could be detected as a methylation change include single-nucleotide polymorphisms (SNPs) and rare single-nucleotide variants including *de novo*. We addressed this potential concern in several ways. First, to the extent possible, we removed probes annotated as being influenced by SNPs. Second, our statistical approach searched for DMRs, i.e. multiple CpG probes showing consistent changes in methylation associated with air pollutant exposures, thus, is less likely to be influenced by genetic variation-related methylation changes at a single CpG. Despite our best efforts, it is possible that we were not able to completely remove the potential effects of genetic variation on DNAm. Even if some of the DMRs we report are detecting methylation changes resulting from underlying DNA sequence variation, our results are important because they identify biologic processes and genes associated with air pollutant exposure levels. Future studies that include comprehensive genetic variant, methylation, and air pollution exposure measures from the same individuals are needed to examine what role, if any, genetic variation plays in the biologic mechanism for observed methylation–air pollution associations.

The present study has several potential limitations. First, our sample size is small, with only 38 participants, which may have limited our ability to detect all significant changes and associations in our dataset and to identify DMRs in common across pollutants. This is not an issue that is unique to us, however, as there is a general lack of inclusion of sperm sample collection in prospective pregnancy cohorts due to feasibility, and thus, previous larger studies can only provide cross-sectional data from a more accessible data source [40]. Furthermore, due to the study

design, a subgroup of men from the population already has a child with ASD. While it is hard to envision a scenario where the methylation patterns related to air pollution exposure levels in men who have a child with autism would be fundamentally different from men who do not have an autistic child, it is possible. Thus, exploring whether similar changes are also observed in different groups of men is warranted. These sample size and study design challenges need to be weighed against the particularly unique and valuable enriched cohort that was collected. Future general population epidemiology studies would benefit from collecting sperm from men to address these limitations. Second, air pollution measures were generated based on maternal addresses in EARLI. While we did exclude fathers who reported not living with the child's biological mother from our study, it is possible that the air pollution exposure data we collected are not reflective of the place the father spent the most time (i.e. work or somewhere else outside the home). Third, DNA was extracted from semen as opposed to purified mature motile sperm, making it possible that any associations detected may have been diluted by cellular heterogeneity. However, we were still able to detect meaningful associations between air pollution exposure and DNAm in sperm. Finally, the study population was selected for families at increased likelihood of autism, and findings may not represent associations in the general population.

These limitations are balanced by the strength and novelty of our findings. This study was among the first to demonstrate that genome-scale methylation changes in sperm occur at genes enriched for neurodevelopmental processes in response to air pollution exposure. These findings emphasize the need to extend the early life window to incorporate the paternal preconception environment and to encourage future studies to incorporate epigenetic and environmental data into studies seeking to better understand how paternal factors contribute to neurodevelopmental and other health outcomes in the next generation.

Materials and methods

Study cohort

EARLI is described in detail in the study by Newschaffer *et al.* 2012 [25]. Briefly, EARLI is a prospective study that enrolled pregnant mothers with a child already diagnosed with autism. Infant siblings were followed from birth until 3 years of age. EARLI recruited and enrolled participants from four sites across the USA—San Francisco and Sacramento, CA, Baltimore, MD, and Philadelphia, PA. Sperm samples were collected from fathers around the time of enrollment in EARLI, typically in the first or second trimester of pregnancy [23]. Participants were included in the present study if they had genotyping data, sperm methylation data, air pollution data for all four pollutants ($PM_{2.5}$, PM_{10}, NO_2, and O_3), and covariate data and did not report living in a different house than the child's mother ($n = 38$ included in final sample size). Informed consent was obtained from all EARLI participants included in this study, and this study was approved by the Institutional Review Board at Johns Hopkins University.

Air pollution exposure measurements

Exposure assignments for fathers were based on maternal residences recorded prospectively throughout pregnancy. Because of this, fathers were excluded from this analysis if they reported not living in the same home as the mother. Detailed methods on air pollution measurements can be found as previously described [26]. Briefly, residential addresses for study participants were standardized and geo-coded using the TeleAtlas US_Geo_2 database and software. Assignments for each of the four pollutants were derived from the US EPA's Air Quality System data (www.epa.gov/ttn/airs/airsaqs). Weekly air quality data from monitoring stations located within 50 km of each residence were made available for spatial interpolation of ambient pollutant concentrations, which were based on inverse distance-squared weighting of data from up to four closest stations as previously described [26]. Exposure measurements used in the present analyses were calculated as the average of exposures measured within 4 weeks prior to the date of sample collection to capture the last phase of spermatogenesis and maturation that occurred just prior to sample collection.

Sperm collection and DNA extraction

Sperm collection and DNA extraction methods are as previously described [23, 24]. Briefly, semen samples, frozen upon collection, were shipped frozen to the Johns Hopkins Biological Repository where they were stored at −80°C until processing. Genomic DNA was extracted using the QIAgen QIAsymphony automated workstation following the Blood 1000 protocol from the DNA Midi Kit (Cat. number 937255, Qiagen, Valencia, CA).

DNAm measurement

Genome-scale DNAm in sperm was measured as described in the study by Feinberg *et al.* 2015 [23, 24]. Briefly, DNAm was measured using the CHARM assay. Genomic sperm DNA (4 μg) underwent shearing, digestion with McrBC enzyme, purification, and labeling and hybridization to the arrays as previously described [23]. In addition to probes included on the original CHARM method, probes on the CHARM array cover all promoters and microRNA sites. Raw methylation data were previously uploaded to the National Database for Autism Research (NDAR) under study 377.

DNAm data processing and quality assessment

Raw data from the CHARM array were pre-processed using the CHARM package (v.2.8.0) in R (version 3.0.3) as was previously described and used in other studies [23, 24]. In short, the background signal was removed and probe-level DNAm estimates (percentages) were determined by normalization to control probes. After exclusion of background, control, and repetitive probe groups, 3 811 046 total probes per array remained for each sample. We also determined whether the methylation measurement quality differed by air pollutant type and did not observe relationships between air pollutants and measurement of technical variables including CHARM DNA shearing or hybridization date, shearing matching, CHARM gel, or gel location (Supplementary Table S1a–d). DNAm was also measured for a subset of available sperm samples via the Illumina Infinium HumanMethylation450 BeadChip assay (referred to as 450 K) array (Illumina, San Diego, CA). Detailed protocols for DNAm measurement and cross-validation and data quality assessment are previously described [23, 24]. Genomic DNA (1 μg) was processed by the Johns Hopkins University SNP Center using the automated Infinium workflow. Illumina infinium methylation data for the overlapping subjects were processed using the preprocessNoob function from the minfi package (v 1.22.1) in R version 3.4.0. No probes were excluded during preprocessing.

SV analysis

Surrogate variable analysis was performed on percentage methylation estimates as described in the studies by Feinberg *et al.* 2015 and Leek & Storey 2007 [23, 41] to adjust for any batch effects or

other unwanted sources of variation that might otherwise influence DNAm. R version 4.1 was used to run the "sva" R package, as well as for all other downstream analyses unless otherwise specified. The number of SVs present in the data was estimated using the Buja and Eyuboglu algorithm, and SVs were adjusted for in downstream analyses.

Ancestry principal components

DNA from buffy coat, white blood cells, and saliva were processed on the Affymetrix omni5 exome array at the Johns Hopkins University SNP Center for genotyping analysis. Using PLINK v1.90 [42], parents were subset from the EARLI dataset and merged with the 1000 G Phase3 v5 reference keeping only overlapping SNPs and a minor allele frequency filter of ≥0.05. Principal components (PCs) 1–10 were then assembled using smartpca from EIGENSOFT 6.1.4 [43, 44] using 1000G Phase3 v5 as an anchor for ancestry.

Statistical analyses to identify air pollution–associated changes in regional DNAm

Differences in air pollution exposures across demographic variables were assessed using *t*-tests, analysis of variance, Mann–Whitney *U* test, and Spearman rank correlations. Similar tests were performed when assessing associations between estimated SVs (described later) and demographic variables. Methods for identifying regions of CHARM DNAm that were associated with $PM_{2.5}$, PM_{10}, NO_2, or O_3 are described in detail in the study by Feinberg *et al.* 2015 [23, 24]. In brief, DMRs were identified using the "bump hunting" approach that was developed for the CHARM platform as previously described [23, 45]. In our statistical models, DNAm changes were treated as the outcome of interest, with $PM_{2.5}$, PM_{10}, NO_2, or O_3 as the exposure, and we adjusted for SVs ($n=6$ SVs for $PM_{2.5}$, $n=7$ SVs for PM_{10} and O_3, and $n=8$ for NO_2), four ancestry PCs, paternal education, and coast (East versus West). This model was applied to all high-quality probes present on the CHARM array to identify the adjusted linear effect of air pollution exposure on DNAm levels [23, 45]. Regions of differential methylation were identified by smoothing the linear effects of exposure on DNAm and then thresholding the smoothed statistics across all probes with a 99.9th percentile cutoff, as was previously described, to generate a region-level statistic reflecting the area of a DMR [23, 45]. A *P*-value for each DMR was calculated from a genome-wide empirical distribution of null statistics generated from a linear model bootstrapping approach across 10 000 permutations [23, 45]. DMRs were ranked prior to the permutation procedure to identify the threshold that controls the target genome-wide family-wise empirical (FWER), and DMRs with a FWER $P<0.05$ were considered significant [23, 45].

Gene enrichment analysis

Testing for gene enrichment was previously described [23, 24]. Briefly, we used the GOstats R Bioconductor package to test for enrichment of genes within 10 kb of the transcription start site of the gene for DMRs with FWER $P<0.1$. We restricted our analysis to Biological Process ontology germs with at least four members. Genes mapped to top DMRs (FWER $P<0.1$) were compared to all genes on the CHARM array that also have an Entrez ID as the background.

Cross-platform validation

We attempted to validate DMRs identified via CHARM using overlapping genomic coverage on the 450 K array in a partially overlapping set of sperm samples ($n=25$). Linear regression was used to assess the relationship between CpG-site 450 K DNAm and air pollution exposure at CpG sites that were included in DMRs identified via the CHARM array. Models were adjusted for SVs as well as coast, paternal education, and four ancestry PCs. The correlation between estimates across the two platforms (CHARM and 450 K) was determined using Spearman correlation tests.

Acknowledgements

We would like to thank Rakel Tryggvadottir, Birna Berndsen, Roxann Ashworth, and the Johns Hopkins SNP Center at the Genome Resource Core Facility for processing the laboratory samples. We would like to thank Frederick W Lurmann for his assistance in generating the air pollution measures.

Author contributions

Jason I. Feinberg (laboratory studies [lead], statistical analyses [equal], writing—manuscript [shared]), Rose Schrott (statistical analyses [equal], writing—manuscript [shared]), Andrew P. Feinberg (statistical analyses [shared], writing—manuscript [shared], overall design [equal]), Heather E. Volk (statistical analyses [shared], writing—manuscript [shared], overall design [equal]), Christine Ladd-Acosta (statistical analyses [shared], overall design [shared]), M. Daniele Fallin (statistical analyses [shared], EARLI cohort principal investigators [shared], overall design [shared]), Craig J. Newschaffer (EARLI cohort principal investigators [shared]), Irva Hertz-Picciotto (EARLI cohort principal investigators [shared]), and Lisa A. Croen (EARLI cohort principal investigators [shared]).

Supplementary data

Supplementary data is available at *EnvEpig* online.

Conflict of interest statement. Dr Ladd-Acosta reports receiving consulting fees from the University of Iowa for providing expertise on autism spectrum disorder epigenetics outside of this work. No other authors have anything to declare.

Funding

This work was supported by R01ES017646 (Feinberg/Fallin), R01ES016443 (Newschaffer), R24ES030893 (Fallin), R01ES023780 (Volk), R01ES023780-04S01 (Volk), and Autism Speaks grant number 7785 (Volk).

Data availability

The raw methylation data analyzed in the current study were previously uploaded to the NDAR study 377 and are available at the following link: https://nda.nih.gov/study.html?id=377. Air pollution measures were previously uploaded to NDAR collection 2563. Additional data are available from the corresponding author upon reasonable request.

References

1. Oudin A, Frondelius K, Haglund N *et al.* Prenatal exposure to air pollution as a potential risk factor for autism and ADHD. *Environ Int* 2019;**133**:105149.
2. Breton CV, Yao J, Millstein J *et al.* Prenatal air pollution exposures, DNA methyl transferase genotypes, and associations with newborn LINE1 and alu methylation and childhood blood pressure and carotid intima-media thickness in

the children's health study. *Environ Health Perspect* 2016;**124**:1905–12.
3. Fossati S, Valvi D, Martinez D *et al.* Prenatal air pollution exposure and growth and cardio-metabolic risk in preschoolers. *Environ Int* 2020;**138**:105619.
4. Tingskov Pedersen CE, Eliasen AU, Ketzel M *et al.* Prenatal exposure to ambient air pollution is associated with early life immune perturbations. *J Allergy Clin Immunol* 2023;**151**:212–21.
5. Rahman MM, Shu Y-H, Chow T *et al.* Prenatal exposure to air pollution and autism spectrum disorder: sensitive windows of exposure and sex differences. *Environ Health Perspect* 2022;**130**:17008.
6. Volk HE, Lurmann F, Penfold B *et al.* Traffic-related air pollution, particulate matter, and autism. *JAMA Psychiatry* 2013;**70**:71–7.
7. Margolis AE, Cohen JW, Ramphal B *et al.* Prenatal exposure to air pollution and early-life stress effects on hippocampal subregional volumes and associations with visuospatial reasoning. *Biol Psychiatry Glob Open Sci* 2022;**2**:292–300.
8. Liu R, DeSerisy M, Fox NA *et al.* Prenatal exposure to air pollution and maternal stress predict infant individual differences in reactivity and regulation and socioemotional development. *J Child Psychol Psychiatry* 2022;**63**:1359–67.
9. Allen JL, Oberdorster G, Morris-Schaffer K *et al.* Developmental neurotoxicity of inhaled ambient ultrafine particle air pollution: parallels with neuropathological and behavioral features of autism and other neurodevelopmental disorders. *Neurotoxicology* 2017;**59**:140–54.
10. Rubes J, Selevan SG, Evenson DP *et al.* Episodic air pollution is associated with increased DNA fragmentation in human sperm without other changes in semen quality. *Hum Reprod* 2005;**20**:2776–83.
11. Cheng Y, Tang Q, Lu Y *et al.* Semen quality and sperm DNA methylation in relation to long-term exposure to air pollution in fertile men: a cross-sectional study. *Environ Pollut* 2022;**300**:118994.
12. Wu W, Chen Y, Cheng Y *et al.* Association between ambient particulate matter exposure and semen quality in fertile men. *Environ Health* 2022;**21**:16.
13. Zhang J, Cai Z, Ma C *et al.* Impacts of outdoor air pollution on human semen quality: a meta-analysis and systematic review. *Biomed Res Int* 2020;**2020**:7528901.
14. Schuller A, Bellini C, Jenkins T *et al.* Simulated wildfire smoke significantly alters sperm DNA methylation patterns in a murine model. *Toxics* 2021;**9**:199.
15. Bernhardt L, Dittrich M, Prell A *et al.* Age-related methylation changes in the human sperm epigenome. *Aging* 2023;**15**:1257–78.
16. Keyhan S, Burke E, Schrott R *et al.* Male obesity impacts DNA methylation reprogramming in sperm. *Clin Epigenetics* 2021;**13**:17.
17. Donkin I, Versteyhe S, Ingerslev L *et al.* Obesity and bariatric surgery drive epigenetic variation of spermatozoa in humans. *Cell Metab* 2016;**23**:369–78.
18. Soubry A, Murphy SK, Vansant G *et al.* Opposing epigenetic signatures in human sperm by intake of fast food versus healthy food. *Front Endocrinol* 2021;**12**:625204.
19. Schrott R, Rajavel M, Acharya K *et al.* Sperm DNA methylation altered by THC and nicotine: vulnerability of neurodevelopmental genes with bivalent chromatin. *Sci Rep* 2020;**10**:16022.
20. Murphy SK, Itchon-Ramos N, Visco Z *et al.* Cannabinoid exposure and altered DNA methylation in rat and human sperm. *Epigenetics* 2018;**13**:1208–21.
21. Schrott R, Murphy SK, Modliszewski JL *et al.* Refraining from use diminishes cannabis-associated epigenetic changes in human sperm. *Environ Epigenetics* 2021;**7**:dvab009.
22. Greeson KW, Fowler KL, Estave PM *et al.* Detrimental effects of flame retardant, PBB153, exposure on sperm and future generations. *Sci Rep* 2020;**10**:8567.
23. Feinberg JI, Bakulski KM, Jaffe AE *et al.* Paternal sperm DNA methylation associated with early signs of autism risk in an autism-enriched cohort. *Int J Epidemiol* 2015;**44**:1199–210.
24. Feinberg JI, Schrott R, Ladd-Acosta C *et al.* Epigenetic changes in sperm are associated with paternal and child quantitative autistic traits in an autism-enriched cohort. *Mol Psychiatry* 2023:1–11.
25. Newschaffer CJ, Croen LA, Fallin MD *et al.* Infant siblings and the investigation of autism risk factors. *J Neurodev Disord* 2012;**4**:7.
26. Ladd-Acosta C, Feinberg JI, Brown SC *et al.* Epigenetic marks of prenatal air pollution exposure found in multiple tissues relevant for child health. *Environ Int* 2019;**126**:363–76.
27. Yu W, Zhang L, Wei Q *et al.* O6-Methylguanine-DNA Methyltransferase (MGMT): challenges and new opportunities in glioma chemotherapy. *Front Oncol* 2020;**9**:1547.
28. Bernard BJ, Nigam N, Burkitt K *et al.* SMYD3: a regulator of epigenetic and signaling pathways in cancer. *Clin Epigenetics* 2021;**13**:45.
29. Yoshihara M, Hara S, Tsujikawa M *et al.* Restricted presence of POU6F2 in human corneal endothelial cells uncovered by extension of the promoter-level expression atlas. *EBioMedicine* 2017;**25**:175–86.
30. Sudhof TC. Neuroligins and neurexins link synaptic function to cognitive disease. *Nature* 2008;**455**:903–11.
31. Bill BR, Lowe JK, Dybuncio CT *et al.* Orchestration of neurodevelopmental programs by RBFOX1: implications for autism spectrum disorder. *Int Rev Neurobiol* 2013;**113**:251–67.
32. Salatino-Oliveira A, Genro JP, Polanczyk G *et al.* Cadherin-13 gene is associated with hyperactive/impulsive symptoms in attention/deficit hyperactivity disorder. *Am J Med Genet B Neuropsychiatr Genet* 2015;**168B**:162–9.
33. Kim KS, Park S. Impact of lung-related polygenic risk scores on chronic obstructive pulmonary disease risk and their interaction with w-3 fatty acid intake in middle-aged and elderly individuals. *Nutrients* 2023;**15**:3062.
34. Pedersen RTA, Snoberger A, Pyrpassopoulos S *et al.* Endocytic myosin-1 is a force-insensitive, power-generating motor. *J Cell Biol* 2023;**222**:e202303095.
35. Liu YJ, Yin S-Y, Zeng S-H *et al.* Prognostic value of LHFPL tetraspan subfamily member 6 (LHFPL6) in gastric cancer: a study based on bioinformatics analysis and experimental validation. *Pharmgenomics Pers Med* 2021;**14**:1483–504.
36. Abd El Gayed EM, Rizk MS, Ramadan AN *et al.* mRNA expression of the CUB and Sushi Multiple Domains 1 (CSMD1) and its serum protein level as predictors for psychosis in the familial high-risk children and young adults. *ACS Omega* 2021;**6**:24128–38.
37. Starling AP, Wood C, Liu C *et al.* Ambient air pollution during pregnancy and DNA methylation in umbilical cord blood, with potential mediation of associations with infant adiposity: the healthy start study. *Environ Res* 2022;**214**:113881.
38. Urdinguio RG, Sanchez-Mut JV, Esteller M. Epigenetic mechanisms in neurological diseases: genes, syndromes, and therapies. *Lancet Neurol* 2009;**8**:1056–72.
39. Cukier HN, Dueker ND, Slifer SH *et al.* Exome sequencing of extended families with autism reveals genes shared across neurodevelopmental and neuropsychiatric disorders. *Mol Autism* 2014;**5**:1.

40. Zhang Y, Wei J, Liu C *et al.* Association between ambient PM(1) and semen quality: a cross-sectional study of 27,854 men in China. *Environ Int* 2023;**175**:107919.
41. Leek JT, Storey JD, Gibson G. Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS Genet* 2007;**3**:1724–35.
42. Chang CC, Chow CC, Tellier LC *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 2015;**4**:7.
43. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. *PLoS Genet* 2006;**2**:e190.
44. Price AL, Patterson NJ, Plenge RM *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 2006;**38**:904–9.
45. Jaffe AE, Murakami P, Lee H *et al.* Bump hunting to identify differentially methylated regions in epigenetic epidemiology studies. *Int J Epidemiol* 2012;**41**:200–9.

Environmental Epigenetics, 2024, **10(1)**, 1–14
DOI: https://doi.org/10.1093/eep/dvae003
Advance Access Publication 8 February 2024
Research Article
Received 9 November 2023; revised 10 January 2024; accepted 2 February 2024