

Supplementary Material for:

Prenatal exposures and DNA methylation in newborns: A Pilot Study in Durban, South Africa

Authors and Affiliations:

Jaclyn M. Goodrich^a, Poovendhree Reddy^b, Rajen N. Naidoo^c, Kareshma Asharam^c, Stuart Batterman^a, Dana C. Dolinoy^{a,d}

^a Department of Environmental Health Sciences, University of Michigan School of Public Health, Ann Arbor, MI USA

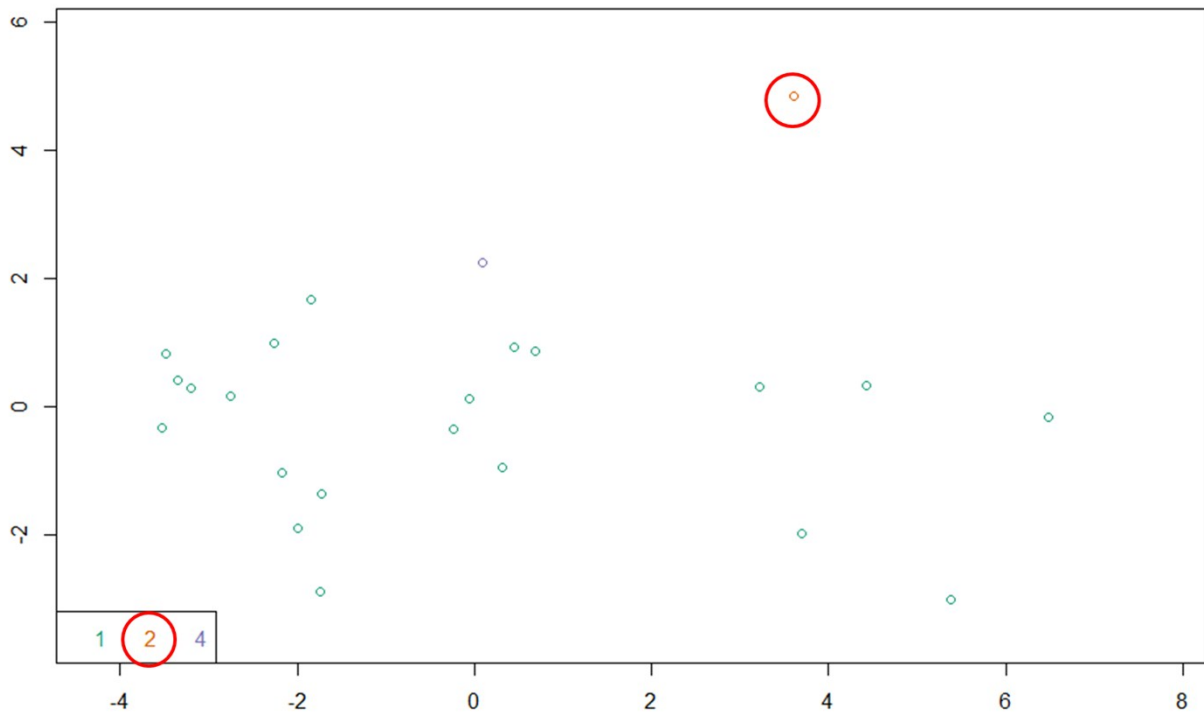
^b Department of Community Health Studies, Durban University of Technology, Durban, South Africa

^c Discipline of Occupational and Environmental Health, School of Nursing and Public Health, University of KwaZulu-Natal, Durban, South Africa

^d Department of Nutritional Sciences, University of Michigan School of Public Health, Ann Arbor, MI USA

SUPPLEMENTAL FIGURE

Supplemental Figure 1. Cluster Analysis of the Most Variable CpG Sites by Race. Subjects were clustered using a multi-dimensional scaling (MDS) algorithm based on the 1000 most variable CpG sites from the 450K analysis after pre-processing of the data (removal of cross-reactive, SNP-containing, and X and Y probes, background and dye bias correction, BMIQ, and batch adjustment). Subjects are colored by self-reported race (green 1=African, orange 2=White, purple 4=Coloured). Since the white subject (circled in red) clustered away from the rest and race is known to influence the epigenome, analyses were run without this subject.



SUPPLEMENTAL TABLES

Supplemental Table I. Top Differentially Methylated Sites by Air Pollution Exposure Region (High, South Durban vs. Low, North Durban). Results are shown from the model adjusting for gestational age, gender, and maternal HIV status (n=22).

Top sites were selected that had $p < 0.0001$ in at least 2 models (by region of Durban, by region in model also adjusting for HIV, and by NO_x levels).

Probe ID	Gene Symbol	Chr	Position	log Fold Change*	P-Value	adj. P-Value**	Relative to CpG Island	Annotation
cg00688840		14	106363051	1.36	2.65E-07	0.11	Open Sea	
cg02092102		7	128555896	-1.22	1.49E-05	0.83	Island	
cg25406779	<i>CNTD2</i>	19	40732712	-0.97	1.10E-05	0.83	South Shore	TSS 200
cg27343801	<i>NOVA2</i>	19	46444017	-1.31	1.66E-05	0.83	Island	Body
cg20791839	<i>SELK</i>	3	53926047	-0.76	7.96E-06	0.83	Island	TSS 200
cg13436843	<i>TRIM31</i>	6	30071601	-0.93	2.62E-05	0.83	South Shore	Body
cg24451081	<i>F7</i>	13	113770401	0.69	3.59E-05	0.83	North Shore	Body
cg27353629		8	144309591	-0.97	3.30E-05	0.83	North Shelf	
cg18739405	<i>KIAA1731</i>	11	93394616	-0.54	4.68E-05	0.83	North Shore	TSS 200
cg24324441	<i>MFSD1</i>	3	158519002	-0.91	4.30E-05	0.83	North Shore	TSS 1500
cg18551877	<i>GLYATL2</i>	11	58612087	1.97	4.85E-05	0.83	Open Sea	TSS 200
cg20905180	<i>MTMR9</i>	8	11149412	1.48	4.95E-05	0.83	Open Sea	Body
cg13388615		2	242878809	-1.52	1.77E-05	0.83	Island	
cg25398924	<i>DLGAP2</i>	8	1460718	1.04	7.55E-05	0.89	Open Sea	5' UTR
cg06750897	<i>PBX1</i>	1	164545553	-0.86	6.48E-05	0.89	Island	Body
cg24051234		10	111912029	-3.95	5.52E-05	0.86	Open Sea	
cg13710375	<i>ZNF845</i>	19	53836837	-1.13	9.64E-05	0.89	Island	TSS 200
cg06889422	<i>GGT5</i>	22	24627294	0.61	1.01E-04	0.89	Open Sea	Body

cg15324756	<i>COL4A2</i>	13	110961142	-0.69	4.73E-05	0.83	Island	Body
cg03640051	<i>ATP8A1</i>	4	42657689	0.86	7.69E-05	0.89	North Shore	

*Log fold change is the difference in M-values (\log_2 methylation) of south Durban compared to north Durban

**By Benjamini Hochberg method, based on 433,782 models.

Supplemental Table II. Top Differentially Methylated Sites by Maternal HIV Status. Results are shown from the model adjusting for gestational age, gender, and region of Durban (n=22). Top sites were selected by HIV status that had $p < 0.0001$ in models adjusting for gestational age and gender (with or without adjustment for air pollution exposure group).

Probe ID	Gene Symbol	Chr	Position	log Fold Change*	P-Value	adj. P-Value**	Relative to CpG Island	Annotation
cg17820299	<i>POLDIP3</i>	22	43011819	-1.03	1.67E-05	0.46	South Shore	TSS 1500
cg20402079	<i>RHOBTB3</i>	5	95066792	-1.11	1.99E-05	0.46	North Shore	TSS 200
cg12319534		4	54959841	-0.83	1.13E-05	0.46	South Shore	
cg22758680	<i>ZNF398</i>	7	148846616	0.68	3.42E-05	0.46	South Shore	5' UTR
cg18252731	<i>FAM174B</i>	15	93198700	-1.07	3.15E-05	0.46	Island	1st Exon
cg11114853	<i>DTYMK</i>	2	242626719	-1.18	3.60E-05	0.46	Island	TSS 1500
cg00939727		2	134935506	-0.64	7.35E-05	0.46	Open Sea	
cg13594151	<i>ZNF44</i>	19	12405719	-0.94	4.57E-05	0.46	South Shore	TSS 200
cg17322000	<i>TOP1MT</i>	8	144408126	-0.92	7.33E-05	0.46	Open Sea	Body
cg24001801		7	9057127	-0.91	3.65E-05	0.46	Open Sea	
cg26539615	<i>TULP2</i>	19	49402007	0.65	8.82E-05	0.46	North Shore	TSS 200
ch.19.1200890F		19	37240674	-0.97	8.14E-05	0.46	Open Sea	

*Log fold change is the difference in M-values (\log_2 methylation) of HIV positive exposure compared to HIV negative

**By Benjamini Hochberg method, based on 433,782 models.

Supplemental Table III. Top Gene Ontology or Function Pathways by Air Pollution Exposure Analysis. Full results from

LRpath gene set enrichment testing are displayed for concepts significant at a 10% FDR level.

Name	Concept Type	# Genes in Concept	Coefficient	Odds Ratio	P-Value	FDR q	Direction [†]
Ascorbate and aldarate metabolism	KEGG	19	-1.28	3.49E-04	1.54E-05	0.003	↑
oxygen transport	GOBP	11	-1.68	2.90E-05	9.58E-07	0.005	↓
Pentose and glucuronate interconversions	KEGG	24	-1.17	7.10E-04	6.50E-05	0.007	↓
Drug metabolism - cytochrome P450	KEGG	62	-0.94	2.92E-03	1.80E-04	0.013	↓
Drug metabolism - other enzymes	KEGG	39	-1.00	1.99E-03	3.17E-04	0.017	↓
oxygen binding	GOMF	32	-1.33	2.55E-04	2.04E-05	0.019	↓
Retinol metabolism	KEGG	54	-0.91	3.42E-03	5.58E-04	0.024	↓
Other types of O-glycan biosynthesis	KEGG	37	-0.97	2.34E-03	6.74E-04	0.024	↓
gas transport	GOBP	15	-1.50	8.70E-05	1.51E-05	0.025	↓
detection of chemical stimulus involved in sensory perception	GOBP	357	0.68	70.29	1.54E-05	0.025	↑
RNA splicing, via endonucleolytic cleavage and ligation	GOBP	10	-1.61	4.65E-05	2.54E-05	0.029	↓
detection of chemical stimulus	GOBP	388	0.64	54.93	3.90E-05	0.029	↑
cellular glucuronidation	GOBP	13	-1.51	8.34E-05	4.08E-05	0.029	↓
uronic acid metabolic process	GOBP	14	-1.48	9.83E-05	4.72E-05	0.029	↓
glucuronate metabolic process	GOBP	14	-1.48	9.83E-05	4.72E-05	0.029	↓
Metabolism of xenobiotics by cytochrome P450	KEGG	63	-0.86	4.81E-03	9.74E-04	0.030	↓
sensory perception of chemical stimulus	GOBP	403	0.63	49.35	5.74E-05	0.031	↑
glucuronosyltransferase activity	GOMF	22	-1.39	1.78E-04	7.36E-05	0.031	↓
olfactory receptor activity	GOMF	320	0.65	55.65	1.02E-04	0.031	↑
detection of stimulus involved in sensory perception	GOBP	396	0.63	48.97	6.59E-05	0.032	↑

Name (Table I continued)	Concept Type	# Genes in Concept	Coefficient	Odds Ratio	P-Value	FDR q	Direction
negative regulation of intrinsic apoptotic signaling pathway in response to DNA damage	GOBP	23	1.06	744.26	7.37E-05	0.033	↑
detection of chemical stimulus involved in sensory perception of smell	GOBP	320	0.64	54.31	1.16E-04	0.047	↑
Porphyrin and chlorophyll metabolism	KEGG	33	-0.94	2.85E-03	1.92E-03	0.051	↓
diacylglycerol binding	GOMF	10	1.15	1239.89	2.71E-04	0.063	↑
Starch and sucrose metabolism	KEGG	38	-0.90	3.74E-03	2.68E-03	0.063	↓
regulation of collateral sprouting	GOBP	12	-1.48	1.03E-04	1.76E-04	0.067	↓
Olfactory transduction	KEGG	337	0.46	17.86	3.34E-03	0.071	↑
negative regulation of fibroblast proliferation	GOBP	23	1.02	582.40	2.41E-04	0.085	↑
negative regulation of response to DNA damage stimulus	GOBP	37	0.95	358.40	2.86E-04	0.086	↑
regulation of intrinsic apoptotic signaling pathway in response to DNA damage	GOBP	30	0.98	436.26	2.88E-04	0.086	↑
sensory perception of smell	GOBP	344	0.60	42.06	2.97E-04	0.086	↑
purinergic nucleotide receptor activity	GOMF	14	1.06	742.10	6.33E-04	0.098	↑
nucleotide receptor activity	GOMF	14	1.06	742.10	6.33E-04	0.098	↑
negative regulation of adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains	GOBP	13	-1.42	1.43E-04	3.64E-04	0.100	↓

† Up direction signifies that most genes in the concept were hypermethylated in the south Durban group compared with the north Durban group. Vice versa for the down direction.

Supplemental Table IV. Top Gene Ontology or Function Pathways by Maternal HIV Status. Full results from LRpath gene set

enrichment testing are displayed for concepts significant at a 10% FDR level.

Name	Concept Type	# Genes in Concept	Coefficient	Odds Ratio	P-Value	FDR q	Direction [†]	Viral Related [‡]
olfactory receptor activity	GOMF	320	-1.26	3.90E-04	4.14E-39	3.83E-36	↓	
detection of chemical stimulus involved in sensory perception	GOBP	357	-1.23	4.82E-04	1.54E-39	7.59E-36	↓	
detection of chemical stimulus involved in sensory perception of smell	GOBP	320	-1.24	4.61E-04	1.10E-38	1.91E-35	↓	
sensory perception of chemical stimulus	GOBP	403	-1.19	6.17E-04	1.16E-38	1.91E-35	↓	
detection of stimulus involved in sensory perception	GOBP	396	-1.19	6.27E-04	2.59E-38	3.19E-35	↓	
sensory perception of smell	GOBP	344	-1.21	5.57E-04	8.28E-38	8.15E-35	↓	
detection of chemical stimulus	GOBP	388	-1.18	6.73E-04	1.97E-37	1.62E-34	↓	
Olfactory transduction	KEGG	337	-1.43	1.36E-04	6.24E-33	1.33E-30	↓	
detection of stimulus	GOBP	553	-1.02	1.72E-03	1.50E-32	1.06E-29	↓	
sensory perception	GOBP	758	-0.87	4.42E-03	8.62E-27	5.31E-24	↓	
G-protein coupled receptor activity	GOMF	705	-0.86	4.92E-03	2.12E-24	9.80E-22	↓	
Ascorbate and aldarate metabolism	KEGG	19	-1.18	6.36E-04	3.12E-08	3.32E-06	↓	
Retinol metabolism	KEGG	54	-0.95	2.78E-03	1.08E-07	7.46E-06	↓	
Pentose and glucuronate interconversions	KEGG	24	-1.10	1.10E-03	1.40E-07	7.46E-06	↓	
Metabolism of xenobiotics by cytochrome P450	KEGG	63	-0.89	4.06E-03	4.25E-07	1.81E-05	↓	
glucuronosyltransferase activity	GOMF	22	-1.09	1.15E-03	5.92E-08	1.83E-05	↓	
Starch and sucrose metabolism	KEGG	38	-0.95	2.65E-03	1.12E-06	3.49E-05	↓	
Drug metabolism - cytochrome P450	KEGG	62	-0.86	4.67E-03	1.23E-06	3.49E-05	↓	
Steroid hormone biosynthesis	KEGG	47	-0.91	3.46E-03	1.31E-06	3.49E-05	↓	
Porphyrin and chlorophyll metabolism	KEGG	33	-0.95	2.66E-03	3.13E-06	7.42E-05	↓	
Other types of O-glycan biosynthesis	KEGG	37	-0.91	3.59E-03	8.72E-06	1.86E-04	↓	
Drug metabolism - other enzymes	KEGG	39	-0.84	5.49E-03	6.90E-05	0.001	↓	
cellular glucuronidation	GOBP	13	-1.10	1.07E-03	4.49E-06	0.002	↓	

Name (Table II continued)	Concept Type	# Genes in Concept	Coefficient	Odds Ratio	P-Value	FDR q	Direction	Viral Related
uronic acid metabolic process	GOBP	14	-1.08	1.20E-03	5.13E-06	0.002	↓	
glucuronate metabolic process	GOBP	14	-1.08	1.20E-03	5.13E-06	0.002	↓	
Taste transduction	KEGG	48	-0.68	1.48E-02	2.53E-03	0.045	↓	
translation initiation factor binding	GOMF	20	0.82	167.79	2.24E-04	0.052	↑	
protein self-association	GOMF	36	0.75	102.61	3.33E-04	0.052	↑	
myosin binding	GOMF	34	0.80	146.00	3.63E-04	0.052	↑	
ubiquitin binding	GOMF	68	0.67	66.19	4.71E-04	0.052	↑	
small conjugating protein binding	GOMF	80	0.66	61.78	4.79E-04	0.052	↑	
endonuclease activity, active with either ribo- or deoxyribonucleic acids and producing 3'-phosphomonoesters	GOMF	14	-0.97	2.34E-03	5.05E-04	0.052	↓	
negative regulation of fibroblast proliferation	GOBP	23	0.77	121.80	2.00E-04	0.069	↑	
negative regulation of intrinsic apoptotic signaling pathway in response to DNA damage	GOBP	23	0.81	153.83	2.14E-04	0.069	↑	
negative regulation of viral process	GOBP	71	0.77	121.38	2.52E-04	0.069	↑	*
regulation of intrinsic apoptotic signaling pathway in response to DNA damage	GOBP	30	0.77	121.95	2.52E-04	0.069	↑	
negative regulation of multi-organism process	GOBP	101	0.70	75.72	2.63E-04	0.069	↑	
negative regulation of nitrogen compound metabolic process	GOBP	990	0.36	9.66	2.66E-04	0.069	↑	
negative regulation of nucleobase-containing compound metabolic process	GOBP	970	0.37	9.74	2.79E-04	0.069	↑	
viral release from host cell	GOBP	23	0.86	203.73	2.81E-04	0.069	↑	*
negative regulation of response to DNA damage stimulus	GOBP	37	0.76	110.64	2.87E-04	0.069	↑	
negative regulation of viral release from host cell	GOBP	14	0.93	331.61	2.96E-04	0.069	↑	*
regulation of viral release from host cell	GOBP	21	0.88	238.84	3.06E-04	0.069	↑	*
negative regulation of viral transcription	GOBP	23	0.88	233.50	3.25E-04	0.070	↑	*
regulation of viral process	GOBP	147	0.61	43.10	3.71E-04	0.073	↑	*
negative regulation of macromolecule biosynthetic process	GOBP	983	0.36	9.14	3.72E-04	0.073	↑	

Name (Table II continued)	Concept Type	# Genes in Concept	Coefficient	Odds Ratio	P-Value	FDR q	Direction	Viral Related
protein ubiquitination involved in ubiquitin-dependent protein catabolic process	GOBP	73	0.65	57.40	4.55E-04	0.086	↑	
positive regulation of NF-kappaB transcription factor activity	GOBP	110	0.61	45.54	4.81E-04	0.086	↑	
regulation of viral transcription	GOBP	73	0.64	54.60	4.89E-04	0.086	↑	*
Ubiquitin mediated proteolysis	KEGG	127	0.43	14.92	5.38E-03	0.088	↑	
positive regulation of cellular component movement	GOBP	265	0.51	23.51	5.43E-04	0.092	↑	
regulation of symbiosis, encompassing mutualism through parasitism	GOBP	167	0.56	32.24	5.70E-04	0.092	↑	
inositol metabolic process	GOBP	10	-1.02	1.74E-03	6.07E-04	0.092	↓	
positive regulation of sequence-specific DNA binding transcription factor activity	GOBP	192	0.54	28.38	6.19E-04	0.092	↑	
negative regulation of intrinsic apoptotic signaling pathway	GOBP	65	0.63	49.19	6.29E-04	0.092	↑	
gas transport	GOBP	15	-0.94	2.82E-03	6.34E-04	0.092	↓	
positive regulation of type I interferon production	GOBP	70	0.62	46.89	6.61E-04	0.092	↑	*
regulation of multi-organism process	GOBP	279	0.49	21.11	6.76E-04	0.092	↑	
type I interferon production	GOBP	101	0.59	38.78	7.21E-04	0.092	↑	*
viral process	GOBP	646	0.39	11.01	7.25E-04	0.092	↑	*
regulation of type I interferon production	GOBP	99	0.59	38.92	7.30E-04	0.092	↑	*
regulation of fibroblast proliferation	GOBP	66	0.61	45.03	7.58E-04	0.092	↑	
positive regulation of cell migration	GOBP	251	0.50	21.92	7.65E-04	0.092	↑	
positive regulation of cell motility	GOBP	258	0.49	21.08	8.14E-04	0.095	↑	
multi-organism cellular process	GOBP	654	0.38	10.50	8.63E-04	0.097	↑	
fibroblast proliferation	GOBP	69	0.60	42.06	8.63E-04	0.097	↑	
solute:proton symporter activity	GOMF	12	-0.98	2.29E-03	1.07E-03	0.099	↓	

† Up direction signifies that most genes in the concept were hypermethylated in the maternal HIV positive group compared with the negative group. Vice versa for the down direction.

‡ * Denotes that the concept is involved in response to viral infection.