DNA methylation signatures in the Normative Aging Study:

Epigenome-wide association analyses of air pollution exposure, biological aging, metabolism, and lung function decline

Juan J. Carmona[1,2], Richard T. Barfield[3], Allan C. Just[1,2], Elena Colicino[1], Paolo Testa[1], Pia Pafundi[1], Amar J. Mehta[1,4], Cheng Peng[1], Jun Chen[3], Joel Schwartz[1,2,4], Xihong Lin[3], and Andrea Baccarelli[1,2,4,5,6]

1. Laboratory of Environmental Epigenetics, Department of Environmental Health, Harvard School of Public Health; **2**. Exposure, Epidemiology, & Risk Program, Department of Environmental Health, Harvard School of Public Health; **3.** Department of Biostatistics & Program in Quantitative Genomics, Harvard School of Public Health; **4**. Department of Epidemiology, Harvard School of Public Health; **5**. Dana-Farber/Harvard Cancer Center; **6**. Harvard/Massachusetts General Hospital Center on Genomics, Vulnerable Populations & Health Disparities

Background

Epigenetic modifications may serve as indicators of past toxic exposures and predict future disease risk. We propose to discover and validate novel methylomic biomarkers of air pollution exposure and related phenotypic outcomes of interest. Our understanding about the complex interplay of epigene-environment interactions remains rudimentary, and it often based on high-exposure animal models. This array-based methylomics study employs the Normative Aging Study (NAS) cohort, followed for over 40+ years, to identify key epigenetic pathways in humans; the Illumina HumanMethylation450 BeadChip was used to query the methylation status of ~480K CpG sites across the human genome. These epigenetic marks may aid in the early diagnosis and prevention of air pollution-related diseases and the study of basic biological processes *in vivo*.

Objectives

Collectively, we are looking at changes in DNA methylation with various phenotypes/outcomes:

1.) Fasting blood glucose levels
2.) Biological age
3.) Black carbon (BC) & Lung function decline (FEV1)

Methods

Study design: We analyzed ~46,900 CpG sites with the 10% highest variance in methylation in our cohort. We did a cross sectional analysis,

$$CpG_j = \alpha_j Phenotype_j + X_j\beta_j + \varepsilon_j$$

$$\varepsilon_j \sim N(0,\sigma^2)$$

where the outcome was set as fasting blood glucose, black carbon, or aging, and X specified the covariates to be included in the model. j ranged from 1 to m, where m was the number of CpG sites included (Barfield et al., 2012).

- **Confounders for Fasting Blood Glucose**: Age, BMI, insulin intake, other diabetic medications, and smoking status.
- **Confounders for Aging**: BMI, smoking status, physical activity, educational level, alcohol consumption
- **Confounders for Black Carbon and Lung Function Decline**: Age, BMI, smoking status, height, medication intake, education, and disease status

Results

Table 1 Characteristics of participants in VA Normative Aging Study 1999-2008

	Total number of visits		
	One (n=202)	Two (n=473)	
Characteristics	First visit	First visit	Second visit
Age, years, mean (SD)	74.7 (7.2)	71.5 (6.4)	75.1 (6.4)
Body mass index, kg/m^2			
Mean (SD)	27.9 (4.0)	28.2 (4.1)	27.8 (4.1)
Median (IQR)	27.8 (25.1, 30.0)	27.5 (25.5, 30.1)	27.1 (25.1, 30.0)
Height, cm			
Mean (SD)	173.4 (7.3)	173.8 (6.8)	173.5 (6.9)
Median (IQR)	173.7 (168.1, 177.8)	173.7 (169.4, 178.3)	173.4 (169.2, 178.3)
Weight, kg			
Mean (SD)	84.2 (14.3)	85.2 (14.2)	83.8 (14.3)
Median (IQR)	82.8 (73.0, 93.4)	83.5 (76.7, 92.5)	82.1 (74.4, 91.6)
Waist-hip ratio			
Mean (SD)	0.99 (0.05)[a]	0.99 (0.05)[a]	1.00 (0.05)[a]
Median (IQR)	0.99 (0.96, 1.03)[a]	0.99 (0.96, 1.03)[a]	1.00 (0.97, 1.03)[a]
Race, n (%)			
Others	7 (3.5)	11 (2.3)	11 (2.3)
White	194 (96.0)	456 (96.4)	456 (96.4)
Missing	1 (0.5)	6 (1.3)	6 (1.3)
Years of education			
Mean (SD)	14.9 (2.8)[a]	15.1 (3.0)	15.1 (3.0)
Median (IQR)	14.0 (12.0, 16.0)[a]	15.0 (12.0, 17.0)	15.0 (12.0, 17.0)
≥ 2 drinks per day, n (%)	44 (21.8)	87 (18.4)	89 (18.8)
Smoking status, n (%)			
Current	7 (3.5)	20 (4.2)	18 (3.8)
Former	141 (69.8)	314 (66.4)	317 (67.0)
Never	54 (26.7)	139 (29.4)	138 (29.2)
Cumulative pack years smoked			
Mean (SD)	22.7 (26.8)	20.5 (24.7)	20.6 (24.9)
Median (IQR)	14.9 (0, 37.5)	12 (0, 33)	12 (0, 33)
Antihypertensive medication use, n (%)	125 (61.9)	261 (55.2)	322 (68.1)
Coronary heart disease, n (%)	65 (32.2)	123 (26.0)	160 (33.8)
Diabetes Mellitus, n (%)	43 (21.3)	79 (16.7)	96 (20.3)

[a] n missing = 1; [b] n missing = 5 ; [c] n missing = 10; [d] n missing = 7.

Table 2 Selected phenotypes for participants in VA Normative Aging Study 1999-2008

	Total number of visits		
	One (n=202)	Two (n=473)	
Selected outcomes	First visit	First visit	Second visit
FEV_1, L			
Mean (SD)	2.47 (0.58)[a]	2.54 (0.59)[e]	2.54 (0.62)[j]
Median (IQR)	2.50 (2.16, 2.86)[a]	2.56 (2.17, 2.94)[e]	2.53 (2.17, 2.95)[j]
FVC, L			
Mean (SD)	3.29 (0.65)[a]	3.35 (0.67)[e]	3.41 (0.75)[j]
Median (IQR)	3.34 (2.90, 3.70)[a]	3.30 (2.94, 3.79)[e]	3.40 (2.94, 3.90)[j]
FEV_1/FVC			
Mean (SD)	63.7 (27.4)[b]	71.1 (19.1)[f]	56.7 (31.9)[k]
Median (IQR)	75 (65, 80)[b]	76 (70, 80)[f]	72 (52, 78)[k]
SBP, mmHg			
Mean (SD)	128.9 (17.0)[c]	132.1 (17.3)[g]	124.4 (17.0)[c]
Median (IQR)	127.0 (118.5, 139.0)[c]	131.0 (121.0, 142.0)[g]	123.0 (114.5, 135.0)[c]
DBP, mmHg			
Mean (SD)	73.6 (9.5)[c]	77.5 (9.3)[g]	69.8 (9.7)[c]
Median (IQR)	73 (68, 81)[c]	78 (71, 83)[g]	70 (62, 76)[c]
Fasting blood glucose, mg/dL			
Mean (SD)	107.8 (28.5)	107.7 (27.1)[c]	106.1 (21.9)
Median (IQR)	101 (94, 111)	102 (94, 113)[c]	101 (93, 111)
Hemoglobin A1c, tenths of %			
Mean (SD)	5.7 (0.8)[d]	5.5 (0.6)[h]	5.7 (0.7)[l]
Median (IQR)	5.6 (5.3, 6.0)[d]	5.4 (5.2, 5.8)[h]	5.6 (5.2, 5.8)[l]
MMSE (total from 21 trials, max 29)			
Mean (SD)	26.4 (2.9)[e]	26.8 (1.7)[i]	26.6 (1.9)[m]
Median (IQR)	27.0 (26.0, 28.0)[e]	27.0 (26.0, 28.0)[i]	27 (26, 28)[m]

[a] n missing = 32; [b] n missing = 2 ; [c] n missing = 1; [d] n missing = 85; [e] n missing = 33; [f] n missing = 5; [g] n missing = 3; [h] n missing = 391; [i] n missing = 88; [j] n missing = 21; [k] n missing = 8; [l] n missing = 14 ; [m] n missing = 60.

Table 3a Summary statistics short and long-term pollutants relative to examination date for 202 participants with only one visit

Exposures	Moving average						
	4-hr	24-hr	3-day	7-day	14-day	28-day	1-year
$PM_{2.5}$, μg/m^3							
Mean (SD)	11.8 (7.5)	11.0 (6.4)	10.3 (4.5)	10.3 (3.3)	10.4 (2.8)	10.5 (2.5)	11.2 (0.9)
Median (IQR)	9.5 (6.8, 15.6)	8.9 (6.3, 13.9)	9.4 (7.0, 13.0)	9.7 (7.7, 12.1)	10.0 (8.5, 11.8)	10.1 (8.7, 12.0)	11.2 (10.5, 11.9)
BC, μg/m^3							
Mean (SD)	1.2 (0.8)	0.9 (0.4)	0.7 (0.3)	0.8 (0.2)	0.8 (0.2)	0.8 (0.2)	0.55 (0.17)
Median (IQR)	0.9 (0.6, 1.5)	0.7 (0.5, 1.1)	0.7 (0.5, 1.0)	0.7 (0.5, 0.9)	0.8 (0.6, 0.9)	0.8 (0.7, 0.9)	0.54 (0.43, 0.64)
NO_2, ppb							
Mean (SD)	0.018 (0.014)	0.019 (0.005)	0.019 (0.005)	0.019 (0.004)	0.019 (0.004)	0.019 (0.003)	
Median (IQR)	0.014 (0.011, 0.021)	0.019 (0.016, 0.023)	0.019 (0.015, 0.021)	0.019 (0.016, 0.021)	0.019 (0.017, 0.022)	0.020 (0.017, 0.022)	
O_3, ppb							
Mean (SD)	0.014 (0.010)	0.024 (0.012)	0.024 (0.010)	0.024 (0.008)	0.024 (0.008)	0.024 (0.007)	
Median (IQR)	0.013 (0.007, 0.020)	0.023 (0.015, 0.30)	0.024 (0.017, 0.030)	0.025 (0.018, 0.030)	0.026 (0.018, 0.030)	0.026 (0.019, 0.030)	

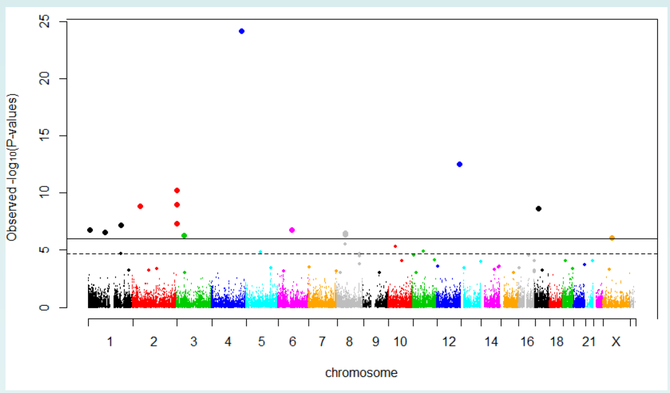
Table 3b Summary statistics short and long-term pollutants relative to examination date for 473 participants at first visit

Exposures	Moving average						
	4-hr	24-hr	3-day	7-day	14-day	28-day	1-year
$PM_{2.5}$, μg/m^3							
Mean (SD)	11.9 (7.1)	11.1 (7.0)	11.0 (5.5)	11.1 (3.5)	11.1 (3.0)	11.1 (2.6)	11.5 (0.9)
Median (IQR)	10.3 (6.5, 15.0)	9.3 (6.8, 13.7)	9.6 (7.4, 13.2)	10.3 (8.9, 12.6)	10.6 (9.0, 12.4)	10.5 (9.2, 12.4)	11.7 (11.0, 12.1)
BC, μg/m^3							
Mean (SD)	1.3 (0.8)	0.9 (0.5)	0.8 (0.3)	0.9 (0.3)	0.8 (0.2)	0.8 (0.2)	0.57 (0.22)
Median (IQR)	1.1 (0.7, 1.8)	0.9 (0.6, 1.3)	0.9 (0.6, 1.0)	0.9 (0.7, 1.0)	0.8 (0.7, 1.0)	0.9 (0.7, 1.0)	0.56 (0.43, 0.68)
NO_2, ppb							
Mean (SD)	0.024 (0.021)	0.022 (0.007)	0.021 (0.005)	0.022 (0.004)	0.021 (0.003)	0.021 (0.003)	
Median (IQR)	0.014 (0.011, 0.021)	0.022 (0.018, 0.026)	0.021 (0.017, 0.024)	0.021 (0.019, 0.024)	0.022 (0.019, 0.023)	0.021 (0.020, 0.023)	
O_3, ppb							
Mean (SD)	0.013 (0.008)	0.022 (0.012)	0.023 (0.011)	0.023 (0.009)	0.023 (0.008)	0.023 (0.008)	
Median (IQR)	0.012 (0.006, 0.018)	0.021 (0.013, 0.29)	0.023 (0.014, 0.030)	0.022 (0.016, 0.030)	0.023 (0.016, 0.030)	0.024 (0.017, 0.030)	

Table 3c Summary statistics short and long-term pollutants relative to examination date for 473 participants at second visit

Exposures	Moving average						
	4-hr	24-hr	3-day	7-day	14-day	28-day	1-year
$PM_{2.5}$, μg/m^3							
Mean (SD)	11.8 (8.6)	10.8 (7.5)	10.3 (6.2)	9.8 (4.0)	9.8 (3.4)	9.8 (2.7)	11.1 (0.9)
Median (IQR)	9.2 (6.6, 13.8)	8.7 (6.2, 12.5)	8.9 (6.6, 12.0)	8.8 (7.6, 11.4)	9.2 (7.6, 11.0)	9.4 (7.9, 10.9)	11.0 (10.4, 11.7)
BC, μg/m^3							
Mean (SD)	1.0 (0.6)	0.8 (0.4)	0.7 (0.3)	0.7 (0.2)	0.7 (0.2)	0.7 (0.2)	0.49 (0.17)
Median (IQR)	0.9 (0.6, 1.3)	0.7 (0.5, 1.3)	0.6 (0.5, 0.9)	0.7 (0.5, 0.9)	0.7 (0.6, 0.8)	0.7 (0.6, 0.8)	0.47 (0.38, 0.57)
NO_2, ppb							
Mean (SD)	0.017 (0.015)	0.019 (0.006)	0.017 (0.005)	0.018 (0.003)	0.018 (0.003)	0.018 (0.003)	
Median (IQR)	0.013 (0.009, 0.020)	0.018 (0.015, 0.021)	0.017 (0.014, 0.020)	0.017 (0.015, 0.020)	0.017 (0.016, 0.020)	0.018 (0.015, 0.019)	
O_3, ppb							
Mean (SD)	0.015 (0.010)	0.024 (0.013)	0.025 (0.010)	0.025 (0.008)	0.025 (0.007)	0.025 (0.007)	
Median (IQR)	0.014 (0.007, 0.022)	0.023 (0.015, 0.31)	0.025 (0.017, 0.032)	0.024 (0.018, 0.031)	0.026 (0.018, 0.031)	0.026 (0.019, 0.031)	

Fig 1 Manhattan Plot for association between methylation and fasting blood glucose level

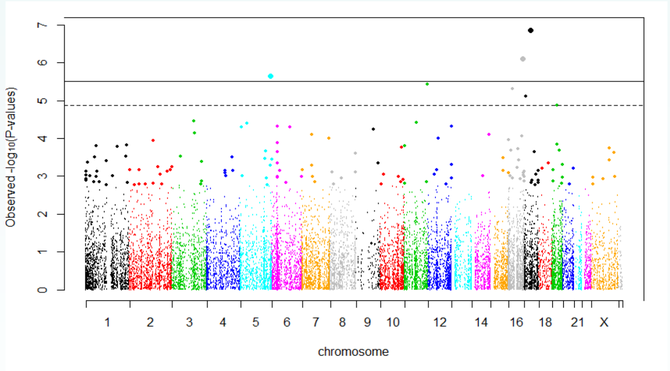


Among the 46,983 CpG sites queried for an association between fasting blood glucose levels and methylation, 23 sites were significant by the BH method.

Selected genes and functions for top 10 CpG hits for the association between methylation and fasting blood glucose level

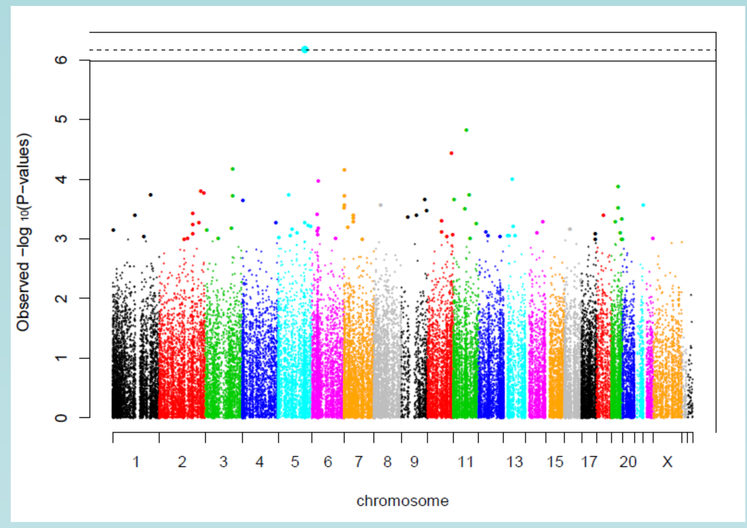
CPG Label	Gene	Function
cg08570521	Peroxisome proliferator activated receptor pathway gene	Potential insulin sensitizer
cg02110031	Gene encoding Ccaat-enhancer-binding protein	Regulates gene in glucose metabolism
cg10415767	Sterol regulatory element-binding transcription factor 1gene	Regulates genes in glucose metabolism
cg24441127	Major histocompatibility complex class 1-related gene	Involve in of autoimmune disease

Fig 2 Manhattan Plot for association between methylation and biological age



3 sites were found significant by the Holm method
7 sites were found significant by BH method

Fig 3 Manhattan plot for association between methylation and 28-day Black Carbon (BC)



No sites were found significant in modeling FEV_1. One site was found significant in relationship to 28-day BC across the CpG sites analyzed.

Discussion

No studies published (to date) have examined associations between short- and long-term exposures to traffic-related air pollution and genome-wide methylation using the 450K.

Here we studied 5-methylcytosine from CpGs on the 450K within the top 10% of highest ratio of variance above the technical replicates in the study, after removing failed samples and probes (wateRmelon pfilter) and background correction.

Fasting blood glucose methylation analysis: One CpG site on chromosome 4 was the most significant hit:

- This CpG site was located at the **peroxisome proliferator-activated receptor pathway** gene (TLL1); components of this receptor pathway are molecular targets for the treatment of diabetes (Celi and Shuldiner, 2002).

Age-associated methylation changes: We found that several CpG sites belong to genes previously implicated in aging biology and related processes:

- **ADAMTS18** encodes a member of the ADAMTS (a disintegrin and metalloproteinase with thrombospondin motifs) protein family, which is a putative tumor suppressor related to nasopharyngeal carcinoma (Li et al., 2010).
- **Glia Cell-Derived Neurotrophic Factor (GDNF)** is a gene subject to epigenetic modifications, contributes to behavioral responses to stress (Uchida et al., 2011).
- Epigenetic process influence the imprinting of **anoctamin 1 (ANO1)**, a calcium activated chloride channel. (Okae et al., 2012)
- **Adenylate cyclase 5 (ADCY5)** is subject to DNA hyper-methylation, which is associated with precancerous stages of lung adenocarciroma (Sato et al., 2013).

Conclusions

Understanding the underlying epigenetic basis of human health and disease outcomes is critical to informing prevention efforts, especially as we reconstruct *past* exposure "signatures" in the epigenome to predict *future* disease risk. Our current study leverages a rich DNA archive to study the association(s) of air pollution, age, lung function decline, and fasting blood glucose on DNA methylation *in vivo*, and our preliminary data suggest that we can identify candidate CpGs in relevant genes that function within basic pathophysiological pathways.

References

- **Barfield R** T, Kilaru V, Smith A K and Conneely K N (2012) CpGassoc: an R function for analysis of DNA methylation microarray data. *Bioinformatics Advance Access*
- **Celi F S** and Shuldiner A R (2002) The role of peroxisome proliferator activated receptor gamma in diabetes and obesity. *Curr. Diab. Rep.* 2:179-85
- **Li Z**, Zhang W, Shao Y, Zhang C, Wu Q, Yang H, Wan X, Zhang J, Guan M, Wan J, Yu B. (2010) High-resolution melting analysis of ADAMTS18 methylation levels in gastric, colorectal and pancreatic cancers. *Med Oncol.* 27(3):998-1004.
- **Okae H**, Hiura H, Nishida Y, Funayama R, Tanaka S, Chiba H, Yaegashi N, Nakayama K, Sasaki H, Arima T. (2012) Re-investigation and RNA sequencing-based identification of genes with placenta-specific imprinted expression. *Hum Mol Genet.* 21(3):548-58.
- **Sato T**, Arai E, Kohno T, Tsuta K, Watanabe S, Soejima K, Betsuyaku T, Kanai Y. (2013) DNA methylation profiles at precancerous stages associated with recurrence of lung adenocarcinoma. *PLoS One.* 8(3):e59444.
- **Uchida S**, Hara K, Kobayashi A, Otsuki K, Yamagata H, Hobara T, Suzuki T, Miyata N, Watanabe Y. (2011) Epigenetic status of Gdnf in the ventral striatum determines susceptibility and adaptation to daily stressful events. *Neuron*. 69(2):359-72.