Title: DNA Methylation Architecture of the *ACE2* gene in Nasal Cells

Authors: Andres Cardenas[1,2*], Sheryl L. Rifas-Shiman[3], Joanne E. Sordillo[3], Dawn L. DeMeo[4], Andrea A Baccarelli[5]; Marie-France Hivert[3,7], MD[2]; Diane R Gold[4,8†]; Emily Oken[3†]

Affiliations: [1]Division of Environmental Health Sciences, School of Public Health, University of California, Berkeley, Berkeley, CA, USA.

[2]Center for Computational Biology, University of California, Berkeley, Berkeley, CA, USA.

[3]Division of Chronic Disease Research Across the Lifecourse, Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, MA, USA.

[4]Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

[5]Department of Environmental Health Sciences, Mailman School of Public Health, Columbia University, New York, NY, USA.

[7]Diabetes Unit, Massachusetts General Hospital, Boston, MA, USA.

[8]Department of Environmental Health, Harvard T.H. Chan School of Public Health, Boston, MA, USA.

[†]These authors contributed equally

[*]Corresponding Author:

Andres Cardenas, PhD, MPH

Division of Environmental Health Sciences, School of Public Health and Center for Computational Biology, University of California, Berkeley, 2121 Berkeley Way, #5121, Berkeley CA 94720

Email: andres.cardenas@berkeley.edu

Funding: This work was supported by the US National Institutes of Health grants R01ES031259, R01HD034568 and UH3OD023286.

ABSTRACT

Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) has led to the global coronavirus disease 2019 (COVID-19) pandemic. SARS-CoV-2 enters cells via angiotensin-Converting Enzyme 2 (ACE2) receptors, highly expressed in nasal epithelium with parallel high infectivity.[1,2] We collected nasal swabs from anterior nares of 547 children, measured DNA methylation (DNAm), and tested differences at 15 *ACE2* CpGs by sex, age, race/ethnicity and epigenetic age. *ACE2* CpGs were differentially methylated by sex with 12 sites having lower DNAm (mean=12.71%) and 3 sites greater DNAm (mean=1.45%) among females. We observed differential DNAm at 5 CpGs for Hispanic females (mean absolute difference=3.22%) and 8 CpGs for Black males (mean absolute difference=1.33%), relative to white participants. Longer DNAm telomere length was associated with greater DNAm at 11 and 13 CpGs among males (mean absolute difference=7.86%) and females (mean absolute difference=8.21%), respectively. Nasal *ACE2* DNAm differences could contribute to our understanding COVID-19 severity and disparities.

The global COVID-19 pandemic caused by SARS-CoV-2 has infected over 23 million individuals and claimed over 810,000 lives worldwide as of August 25, 2020.[3] The transmission of SARS-CoV-2 is thought to mostly occur via respiratory droplets and aerosol, both targeting the respiratory system.[4,5] The SARS-CoV-2 spike (S) glycoprotein allows viral entry into the cell by binding to the extracellular domain of the ACE2 receptor.[6,7] It is thought that ACE2 is necessary for infection by SARS-CoV-2 and that it contributes to disease severity.[8] In a different strain of SARS coronavirus, transgenic mice expressing human ACE2 in airway and other epithelia cells develop a rapidly lethal infection.[9]

The nasal epithelium, in direct contact with the environment, is a primary target of SARS-CoV-2 infection. Both ACE2 expression and SARS-CoV-2 infectivity are higher in nasal epithelium relative to the lungs.[1] The nose is exposed to environmental challenges such as air pollution, cigarette smoke, and allergens that could modulate SARS-CoV-2 infectivity and COVID-19 severity and may explain disparities. For example, it has been hypothesized that exposure to air pollution and cigarette smoke could worsen immune response to SARS-CoV-2.[10,11] This is in part supported by studies demonstrating that ACE2 expression is increased after $PM_{2.5}$ exposure[12,13] and among current smokers.[8,14] Transcriptional regulation via DNAm of ACE2 has been documented in peripheral leukocytes and T cells[15,16] and hypothesized to modulate SARS-CoV-2 infectivity and COVID-19 severity.[17,18]

The *ACE* gene is located in the X-chromosome and shown to escape X-inactivation in females[18] with heterogeneous but significant male-bias in expression in several tissues.[19] However, no study has examined the DNAm landscape of the *ACE2* gene in target nasal cells. Identifying demographic and biological differences could help elucidate disparities and the heterogeneity observed in COVID-19 cases.

RESULTS

A total of 277 males (50.6%) and 270 females (49.4%) provided nasal samples at a mean age of 12.95 years (standard deviation [SD] 0.65; range 11.8 to 15.4 years). A total of 367 participants were white (67.1%), 89 Black (16.3%), 23 Hispanic (4.2%), 17 Asian (3.1%) and 51 (9.3%) more than one race. Mean (SD) for estimated DNAm telomere length and epigenetic age of the skin and blood clock were 7.11 Kb (0.38) and 15.7 years (1.94), **Table S1**.

We used the Illumina EPIC array that estimates DNAm at 16 CpG sites annotated to the *ACE2* gene. These sites cover regions between 1,500-200 and 200-0 bases of the start site, exon, body and 3' UTR regions of the *ACE2* gene. Six CpGs were annotated to a Transcription Factor Binding Site (TFBS) with relatively high evidence from 4-6 experiments extracted from Encyclopedia of DNA Elements (ENCODE) supporting transcription binding at that region. After excluding one CpG due to low detection, DNAm levels of the other 15 CpGs ranged from 2.53% to 95.47% and 14 sites had DNAm levels above 60% with robust significant differences by sex (FDR<0.003), **Table 1 and Figure 1-A**. After adjusting for age, 12 of the 15 CpGs in the *ACE2* gene had significantly lower DNAm levels in females relative to males ranging from -31.56% at cg04013915 to -0.47% at cg18877734, **Table 2**. Three CpGs (cg18458833, cg08559914 and cg05039749) showed greater DNAm in females relative to males; these were also the only sites negatively correlated with DNAm of other CpGs. Most CpGs were strongly positively correlated (r_s>0.5) except for cg05039749, annotated to the body of the gene towards the 3' UTR, showing moderate negative correlations with all other sites, **Figure 1-B**. Given large differences by sex and *ACE2* escaping X-inactivation, we analyzed age associations stratified by sex and observed that chronological age was associated with consistently lower but weak DNAm among females and only 4 CpGs reached unadjusted significance (p<0.05); associations did not

survive false discovery adjustment. No associations with age were observed among males, **Table S2**. However, power was limited by the narrow age range.

Hispanic females had significantly greater DNAm of 6.63% at cg03536816, 1.12% at cg08559914, 4.44% at cg16734967 and 2.21% at cg23232263 while cg10408040 showed a 1.69% lower DNAm relative to white females (FDR<0.05), **Table 2**. A local *ACE2* gene Manhattan plot for the association with Hispanic females with genomic annotation and correlation structure is shown in **Figure 1-C**. Among Black males, significantly lower DNAm at 8 CpG sites was present relative to white males. Specifically, lower DNAm of 0.75% at cg18458833, 0.77% at cg18877734, 1.76% at cg08559914, 3.44% at cg16734967, 1.36% at cg05241917, 1.30% at cg20119767, 0.64% at cg25176872 and 0.60% at cg23232263 for Black relative to white males was observed, adjusting for age (FDR<0.05), **Table 2**. A local *ACE2* gene Manhattan plot for the association among Black males with genomic annotation and correlation structure is presented in **Figure 1-D**. No differences were observed among Asian or participants of more than one race, **Table S2**.

DNAm telomere length (DNAmTL) and the Skin and Blood Epigenetic clock were chosen as markers of epigenetic aging as they performed relatively well against chronological age, **Figure S1**. Longer DNAm telomere length adjusted for age (DNAmTLadjAge) was consistently and strongly associated with greater *ACE2* DNAm for males with a mean increase of 7.86% increased across 11 CpGs. For females a mean of 8.21% greater DNAm across 13 CpGs was observed. Weaker positive relationships were observed for 5 CpGs with epigenetic age acceleration of the skin and blood clock, **Table 3**.

DISCUSSION

In summary, all *ACE2* CpGs tested differed strongly and significantly by sex with most sites showing a decrease in DNAm among females relative to males. In sex stratified analyses, moderate greater DNAm for Hispanic females and consistently lower DNAm among Black males relative to white participants was observed. Longer telomere length adjusted for age was strongly associated with greater *ACE2* DNAm and this was consistent among males and females across most CpGs.

The *ACE2* gene is located in the X-chromosome and escapes X-inactivation in females with potential tissue specificity and heterogenous male expression bias, partially attributed to sex steroids.[20] This heterogeneity has been documented in data from an RNAseq Atlas identifying co-expression of both *ACE2* and the accessory proteases among subsets of respiratory epithelial cells as putative targets of SARS-CoV-2 infection. Specifically, nasal goblet cells and ciliated cells comprised the highest fraction of cells expressing both ACE2 and the viral entry associated protease TMPRSS2.[2,21] The relatively high ACE2 expression in the nasal epithelium its coupled with high infectivity in nasal epithelium with progressively lower expression and infectivity in the bronchial and alveolar regions.[1,22] Increased ACE2 expression has been hypothesized to partially explain differences in COVID-19 severity.[23] For example, lung transcriptome data from patients with comorbid conditions associated with increased risk of severe COVID-19 like hypertension, diabetes, and chronic obstructive lung disease show increased ACE2 expression.[24]

ACE2 expression has been reported to be higher in Asians compared with white and African Americans from type II alveolar cells.[25] However, other reports have found no difference in ACE2 expression by race.[26] Data from lung tissue has shown lower DNAm in females relative to males at cg23232263 and cg16734967 consistent with our finding for females using nasal cells.[27] In addition, a strong negative association between chronological age and DNAm of

cg08559914 from airway epithelial cells was reported. This was not confirmed in our data as associations with age were not observed.[27] However, expression of ACE2 in the lung has been demonstrated to increase with age.[26] In nasal samples, lower ACE2 expression among children relative to older adults has been reported.[28] Age associated expression differences are hypothesized to partially explain the relatively mild COVID-19 documented cases in children. Yet, large racial inequalities in need for hospitalization among children have been reported. Data from the U.S. shows that hospitalizations for Hispanic and Black children are much more common relative to white children.[29] Our results support the hypothesis that *ACE2* hypomethylation in nasal epithelium among black males could lead to increased SARS-CoV-2 infectivity and COVID-19 severity via greater abundance of ACE2 receptors. Results among females need to be cautiously interpreted given ACE2 escape from X-inactivation. *ACE2* genetic variants are rare, not shown to differ by sex or populations.[30,31]

Our study has some limitations. We do not have ACE2 expression which would inform the underlying hypothesis of epigenetic regulation. We hypothesize that the relationship between DNAm, expression and ACE2 receptors is likely complex and depends on the genomic context, cells and life-stage. The absence of age-related DNAm changes could be due to the narrow age range and young profile of our study. However, measures of epigenetic age acceleration might be more sensitive to age-related physiological changes among children as reflected in strong associations with DNAm telomere length. Although most participants in our cohort were white, the relatively large sample allowed us to test differences stratifying by sex and test differences by race/ethnicity.

The epigenome is at the interface of the environment and genomic regulation. Nasal cells which are directly in contact with the environment have been shown to play a key role in SARS-

CoV-2 infection. Our results demonstrate that *ACE2* nasal DNAm reflects differences by sex, race/ethnicity and biological aging. The nasal epigenetic architecture of the *ACE2* gene could contribute to our understanding of COVID-19 and environmental disparities.

ONLINE METHODS

Study Population: We included participants from Project Viva, a prospective pre-birth cohort from eastern Massachusetts recruited between 1999 and 2002 during the initial prenatal visit at Atrius Harvard Vanguard Medical Associates.[32] Eligibility criteria for cohort recruitment included fluency in English, gestational age <22 weeks at the first prenatal visit, and singleton pregnancy. At the early childhood visit, we asked the mother to choose one or more racial/ethnic groups for her child. We created five racial/ethnic groups based on the mother's response including white, Black, Hispanic, Asian, and other (>1 race or other). We used mother's race to fill in missing child race values. Of the total 2128 live births, 547 children were re-contacted during an early-teen in-person visit at a mean age 12.9 years (range; 11.8 to 15.4 y) and provided consent for nasal swab sample collection and epigenetic studies. Mothers provided written informed consent at recruitment and at postpartum follow-up visits. Children provided written assent. The Institutional Review Board of Harvard Pilgrim Health Care reviewed and approved all study protocols.

Nasal DNA methylation analyses: Trained research staff collected nasal swabs from the anterior nares, previously shown to yield respiratory epithelial cells.[33] Nasal swabs were immediately stored in lysis buffer and frozen until processing. We isolated DNA using the Maxwell 16 Buccal Swab LEV DNA Purification Kit following the manufacturer's instructions (Promega, Madison, WI, USA). After DNA extraction from the nasal samples we measured DNA methylation with the Infinium MethylationEPIC BeadChip. All DNAm data were imported into the *R* statistical software and samples were excluded based quality control check as previously described.[34] We preprocessed all of the DNA methylation data using functional normalization to adjust for technical variability[35] and adjusted for probe-type variability using

RCP.[36] Lastly, we used ComBat from the *sva* package to adjust for sample plate (seven 96-well plates) as a technical variable.[37] One CpG site of the *ACE2* gene, cg05748796, was excluded as it failed detection in 143 of the samples (p>0.05). The other 15 CpGs were significantly detected ($p<1x10^{-8}$) across all samples. No SNPs at the CpG, probe or single bases extension were annotated to the 15 *ACE2* CpGs. DNAm methylation was estimated on the β-value scale, from 0-1, reflecting the proportions of DNA molecules that are methylated at the target CpG.

Epigenetic Aging Biomarkers: We calculated two epigenetic aging biomarkers utilizing the Horvath's online calculator (http://dnamage.genetics.ucla.edu/): DNA methylation telomere length (DNAmTL)[38] and the "skin and blood epigenetic clock."[39] We evaluated their performance with scatterplots and Pearson's correlation coefficient (r). We modeled the acceleration of these measures resulting from regressing DNAmTL and the skin and blood epigenetic clock on chronological age. A positive value of DNAmTL adjusted for age (DNAmTLadjAge) would indicate that DNAmTL is longer than expected based on chronological age (thus marker of "younger" cells), while positive residuals from the skin and blood clock would indicate epigenetic age acceleration (thus marker of "older" cells). We modeled the DNAmTLadjAge and epigenetic age acceleration for the skin and blood clock to estimate associations independent of chronological age.

Statistical Analyses: We described our study sample using means, standard deviations (SDs), counts and proportions. We extracted the *ACE2* gene genomic annotation from Bioconductor using the *IlluminaHumanMethylationEPICanno.ilm10b4.hg19* annotation package and calculated means and SDs of DNAm at the selected CpGs in the overall cohort and stratified by sex. We estimated the overall Spearman correlation structure among CpGs and used boxplots to visualize differences in DNAm by sex. We tested differences in DNAm by sex, age,

race/ethnicity, DNAmTLadjAge, and the residuals of the skin and blood clock using robust linear regression models on beta values (0-1) to account for heteroskedasticity or influential points. We estimated differences by sex adjusting for chronological age and estimated differences by age stratified by sex. We estimated differences by race/ethnicity (each other category relative to white, the largest group) adjusted for age and stratified by sex. Differences in epigenetic aging biomarkers were adjusted for race/ethnicity and stratified by sex. We used localized Manhattan plots with the correlation matrix across CpGs to display differences by race/ethnicity using coMET.[40] We report findings as %-DNA methylation differences and adjusted for multiple comparisons for each model independently using a False Discovery Rate of 5% (FDR<0.05). All statistical tests were two sided and analyses were carried out using R, version 4.0.2 (www.r-project.org/).

TABLES

Table 1. Genomic annotation of probes located within the Illumina EPIC array and nasal DNA methylation descriptive statistics for the angiotensin-converting enzyme (*ACE2*) gene located in the X-chromosome and sex differences.

CpG Name	position	Gene Group	TFBS coordinates X-chromosomes	TFBS evidence count	Mean (SD) %-DNA methylation			†Adjusted %-DNAm difference
					Overall	Females	Males	
cg18458833	15621477	TSS1500			95.47 (1.84)	96.28 (0.70)	94.68 (1.84)	1.52
cg21598868	15621167	TSS1500			87.88 (2.51)	81.56 (6.08)	94.05 (2.51)	-12.12
cg18877734	15621084	TSS1500			94.26 (2.22)	94.00 (2.14)	94.51 (2.22)	-0.45
cg03536816	15621042	TSS1500;Body			66.55 (12.88)	56.14 (12.14)	76.69 (12.88)	-21.29
cg08559914	15620240	TSS200			91.80 (4.84)	92.98 (2.88)	90.65 (4.84)	1.84
cg16734967	15620103	1stExon;5'UTR	15619908-15620856	6	77.51 (8.28)	72.46 (9.57)	82.44 (8.28)	-9.76
cg10408040	15619136	5'UTR;ExonBnd			92.76 (1.30)	91.05 (2.75)	94.42 (1.30)	-3.05
cg05241917	15616197	Body	15616057-15616589	4	91.16 (4.48)	89.78 (4.90)	92.49 (4.48)	-2.31
cg20119767	15612969	ExonBnd;Body	15612415-15612981	6	83.72 (5.01)	79.52 (4.90)	87.82 (5.01)	-8.67
cg04013915	15586143	Body	15586103-15586918	6	67.24 (10.23)	51.64 (10.36)	82.44 (10.23)	-31.47
cg16716680	15585368	Body			76.87 (5.76)	66.82 (6.54)	86.68 (5.76)	-20.08
cg25176872	15584988	Body	15582921-15585025	6	89.64 (2.41)	85.20 (4.37)	93.97 (2.41)	-8.49
cg22422976	15584867	Body	15582921-15585025	6	81.54 (2.42)	68.72 (4.97)	94.04 (4.42)	-25.22
cg05039749	15583512	Body			2.53 (0.45)	3.16 (1.47)	1.91 (0.45)	1.00
cg23232263	15579482	3'UTR			90.23 (3.72)	85.52 (5.73)	94.81 (3.72)	-8.96

†Differences in DNAm for females relative to males adjusted by age from a multivariate regression models all results were significant at FDR<0.003

Table 2. Differences in %-DNAm in Black and Hispanic relative to white participants adjusted for age and stratified by sex.

	Females								**Males**						
CpG	**Black**				**Hispanic**				**Black**				**Hispanic**		
	%-DNAm difference	*p*	FDR		%-DNAm difference	*p*	FDR		%-DNAm difference	*p*	FDR		%-DNAm difference	*p*	FDR
cg18458833	-0.16	0.18	0.33		0.11	0.49	0.57		**-0.75**	**0.0011**	**0.005**		0.31	0.49	0.90
cg21598868	1.76	0.03	0.23		2.45	0.11	0.20		-0.47	0.07	0.10		0.34	0.41	0.90
cg18877734	0.47	0.09	0.30		0.60	0.32	0.40		**-0.77**	**0.0036**	**0.011**		0.20	0.73	0.96
cg03536816	0.28	0.88	0.94		**6.63**	**0.0051**	**0.019**		-1.92	0.20	0.22		-0.28	0.92	0.96
cg08559914	-0.71	0.10	0.30		**1.12**	**2.42×10^{-5}**	**0.00036**		**-1.76**	**0.0015**	**0.006**		0.98	0.34	0.90
cg16734967	-1.54	0.30	0.51		**4.44**	**0.0039**	**0.019**		**-3.44**	**4.46×10^{-4}**	**0.005**		1.26	0.35	0.90
cg10408040	0.45	0.18	0.33		**-1.69**	**0.0012**	**0.009**		-0.36	0.04	0.07		0.09	0.65	0.96
cg05241917	0.16	0.77	0.94		0.99	0.21	0.31		**-1.36**	**7.18×10^{-4}**	**0.005**		0.85	0.14	0.90
cg20119767	0.04	0.96	0.96		1.92	0.27	0.37		**-1.30**	**0.02**	**0.045**		0.11	0.91	0.96
cg04013915	1.01	0.52	0.78		3.32	0.05	0.10		-1.66	0.10	0.13		0.70	0.54	0.90
cg16716680	2.07	0.029	0.23		1.17	0.55	0.58		-0.86	0.27	0.27		-1.69	0.24	0.90
cg25176872	1.26	0.08	0.30		2.66	0.024	0.06		**-0.64**	**0.006**	**0.016**		-0.02	0.96	0.96
cg22422976	1.12	0.17	0.33		2.6	0.15	0.25		-0.28	0.20	0.22		0.02	0.89	0.96
cg05039749	0.04	0.86	0.94		0.04	0.90	0.90		0.12	0.10	0.13		-0.08	0.54	0.90
cg23232263	-0.18	0.83	0.94		**2.21**	**0.014**	**0.042**		**-0.60**	**0.016**	**0.03**		0.21	0.54	0.90

Table 3. *Differences in %-DNAm of the *ACE2* gene in nasal cells by estimated age-adjusted DNA methylation telomere length residuals (DNAmTLAdjAge) and epigenetic age acceleration of the skin and blood clock stratified by sex

	DNAmTLadjAge								**Skin & Blood Clock Acceleration**						
CpG	Females				Males				Females				Males		
	%-DNAm difference	*p*	FDR		%-DNAm difference	*P*	FDR		%-DNAm difference	*p*	FDR		%-DNAm difference	*p*	FDR
cg18458833	0.06	0.57	0.61		0.53	0.06	0.08		-0.03	0.16	0.18		-0.01	0.86	0.92
cg21598868	**6.74**	**$1.24x10^{-11}$**	**$2.32x10^{-11}$**		**3.17**	**$6.13x10^{-19}$**	**$1.31x10^{-18}$**		0.39	0.05	0.08		-0.01	0.92	0.92
cg18877734	**2.13**	**$1.06x10^{-10}$**	**$1.77x10^{-10}$**		**2.48**	**$4.43x10^{-8}$**	**$7.39x10^{-8}$**		0.15	0.05	0.08		0.05	0.51	0.92
cg03536816	**24.34**	**$5.06x10^{-124}$**	**$7.59x10^{-123}$**		**28.2**	**$6.50x10^{-53}$**	**$9.74x10^{-52}$**		**0.97**	**0.02**	**0.05**		0.15	0.72	0.92
cg08559914	**4.44**	**$2.14x10^{-25}$**	**$6.43x10^{-25}$**		**7.41**	**$8.60x10^{-27}$**	**$4.30x10^{-26}$**		0.19	0.03	0.06		-0.03	0.83	0.92
cg16734967	**17.32**	**$2.87x10^{-50}$**	**$1.08x10^{-49}$**		**11.68**	**$1.59x10^{-19}$**	**$3.97x10^{-19}$**		**0.92**	**0.015**	**0.05**		0.04	0.86	0.92
cg10408040	0.16	0.71	0.71		0.27	0.16	0.18		0.2	0.02	0.06		-0.05	0.21	0.80
cg05241917	**6.59**	**$9.62x10^{-15}$**	**$2.40\ x10^{-14}$**		**4.18**	**$1.77x10^{-28}$**	**$1.33x10^{-27}$**		**0.37**	**0.004**	**0.03**		-0.08	0.43	0.92
cg20119767	**2.12**	**0.0003**	**0.0004**		**2.48**	**0.0027**	**0.0037**		**0.44**	**0.008**	**0.04**		-0.25	0.07	0.80
cg04013915	**19.83**	**$2.31x10^{-72}$**	**$1.16x10^{-71}$**		**18.2**	**$7.48x10^{-20}$**	**$2.24x10^{-19}$**		**1.15**	**0.004**	**0.03**		0.04	0.88	0.92
cg16716680	**4.15**	**$5.52x10^{-6}$**	**$7.52x10^{-6}$**		0.58	0.52	0.52		0	0.98	0.98		0.18	0.20	0.80
cg25176872	**2.32**	**0.0013**	**0.0015**		**1.08**	**$9.18x10^{-5}$**	**0.00014**		0.22	0.16	0.18		-0.01	0.86	0.92
cg22422976	**5.03**	**$3.67x10^{-13}$**	**$7.86x10^{-13}$**		**1.55**	**$9.13x10^{-13}$**	**$1.71x10^{-12}$**		0.18	0.36	0.38		-0.08	0.15	0.80
cg05039749	**0.72**	**$7.00x10^{-7}$**	**$1.05x10^{-6}$**		-0.08	0.25	0.27		0.09	0.03	0.06		0.01	0.57	0.92
cg23232263	**11.04**	**$4.50Ex10^{-88}$**	**$3.37x10^{-87}$**		**6.05**	**$7.67x10^{-25}$**	**$2.87x10^{-24}$**		0.37	0.08	0.11		0.06	0.38	0.92

*Adjusted for race/ethnicity

Figure 1-A. Boxplots of DNA methylation in nasal cells for the angiotensin-converting enzyme 2 (*ACE2*) gene by sex among 15 CpG sites

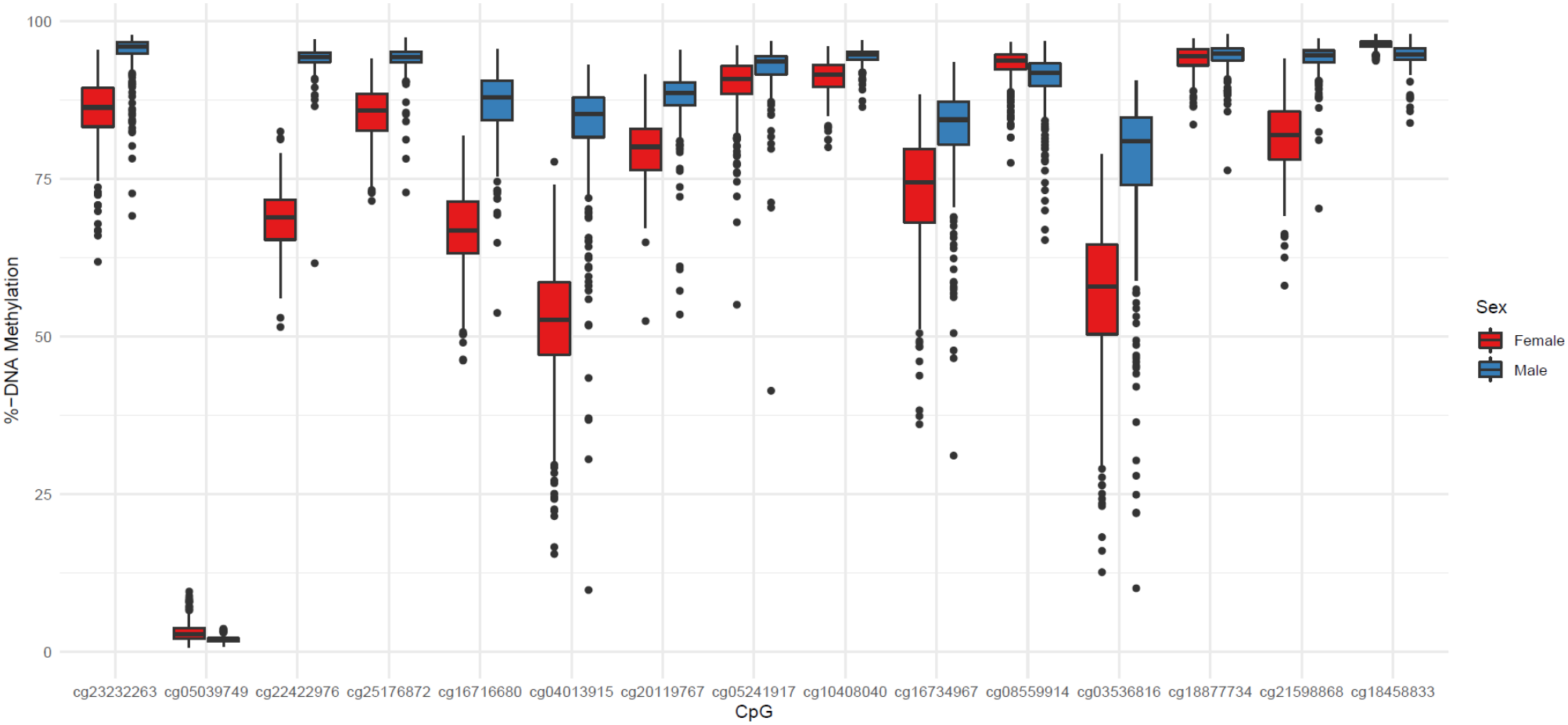


Figure 1-B. Spearman correlation coefficients of nasal cell DNA methylation levels among CpGs found in the angiotensin-converting enzyme (*ACE2*) gene.

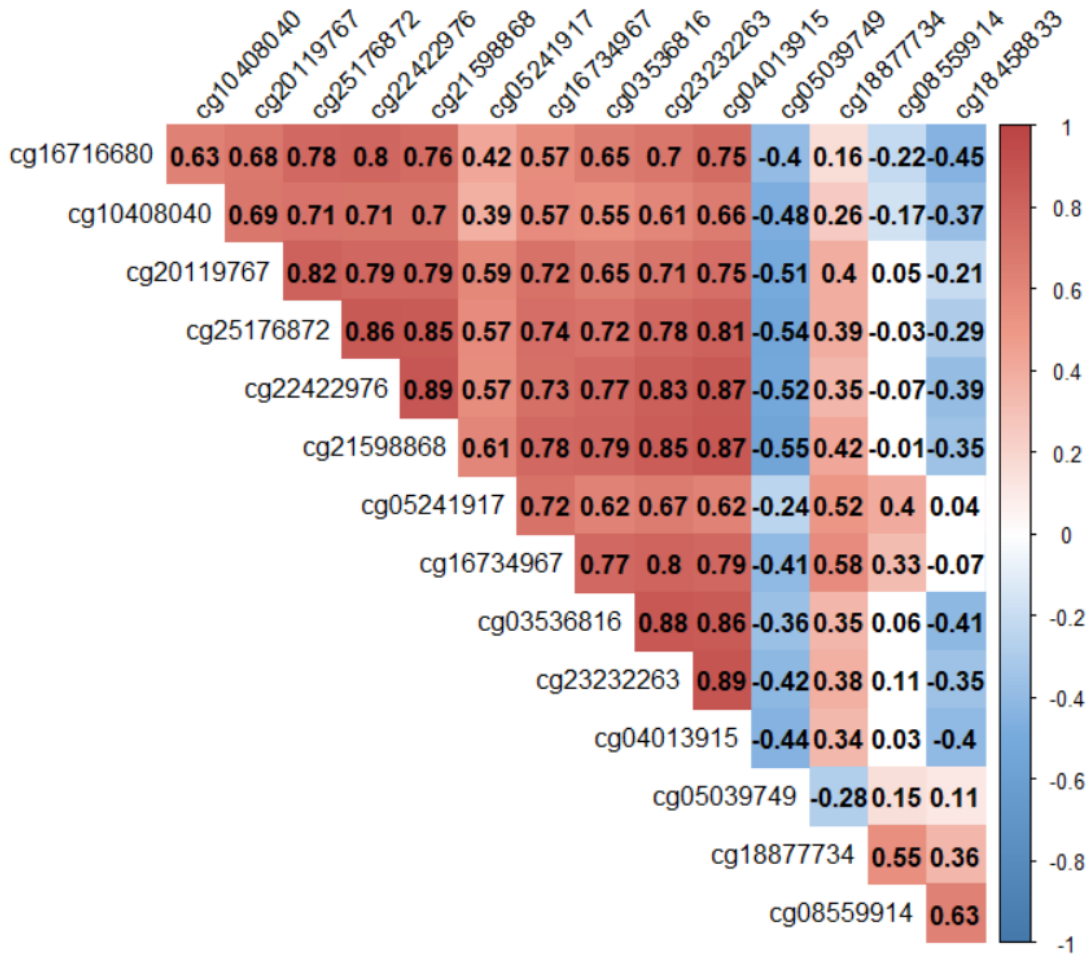


Figure 1-C. Local Manhattan plot for the association of *ACE2* DNA methylation for Hispanic relative to white females. Spearman correlation matrix for CpGs among all females.

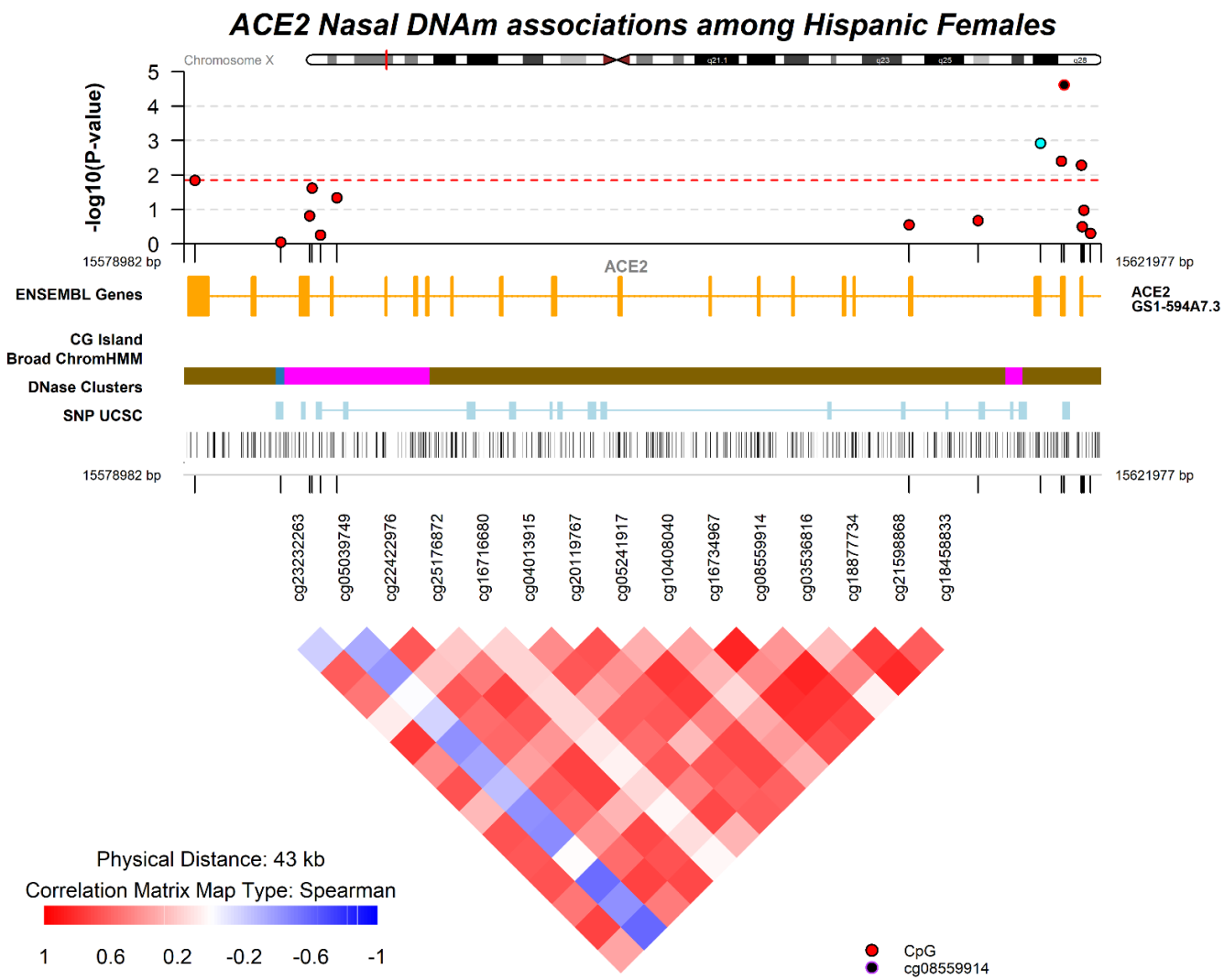
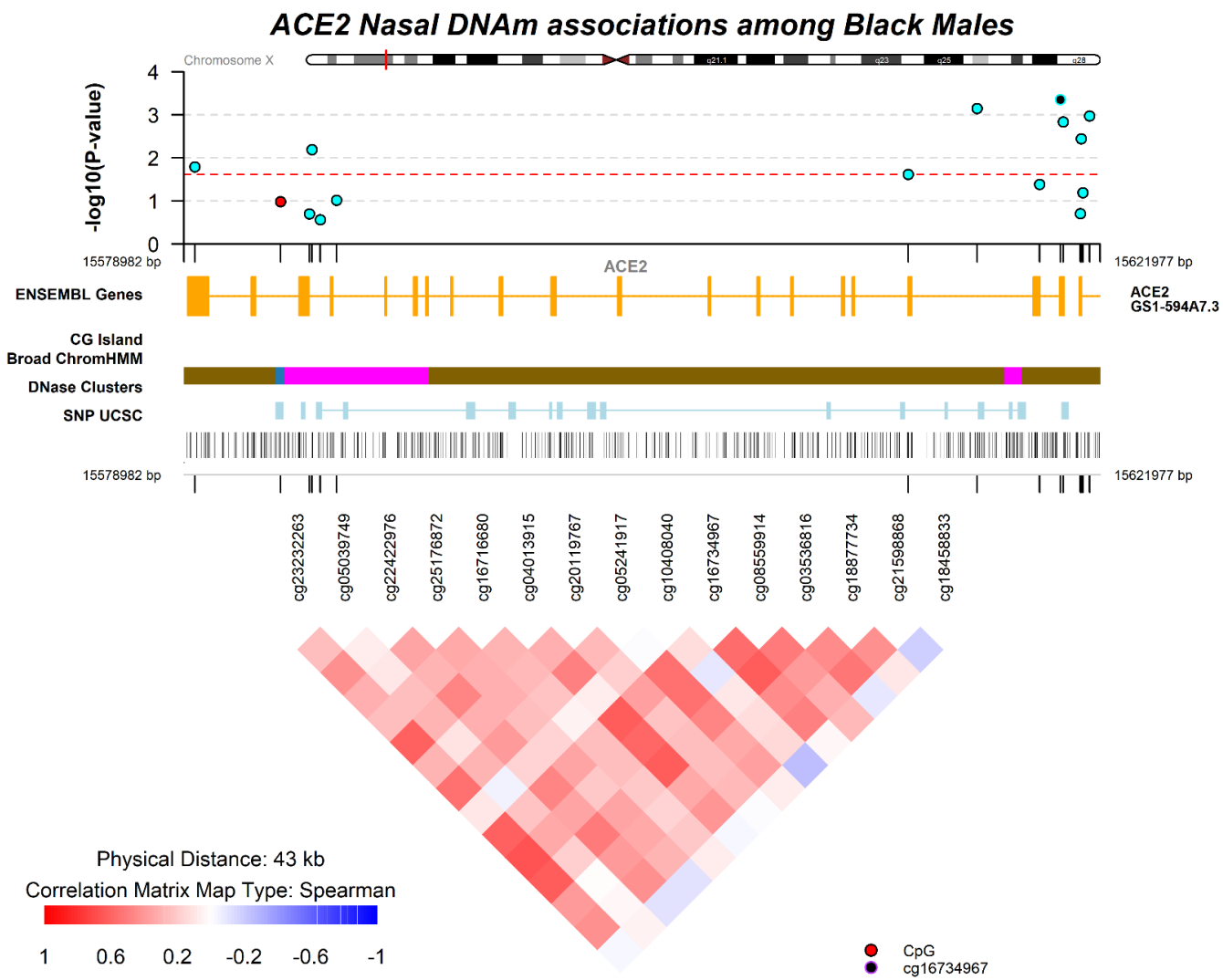


Figure 1-D. Local Manhattan plot for the association of *ACE2* DNA methylation for Black relative to white males. Spearman correlation matrix for CpGs among all males.



Supplementary Information

Table S1. Demographic characteristics among 547 Project Viva participants with nasal DNA methylation measurements.

Participant characteristics	N (%) or mean (SD)
Sex	
Male	277 (50.6%)
Female	270 (49.4%)
Age (years)	12.95 (0.66)
Race/Ethnicity	
White	367 (67.1%)
Black	89 (16.3%)
Hispanic	23 (4.2%)
Asian	17 (3.1%)
More than one race/Other	51 (9.3%)
DNA methylation telomere length (DNAmTL) Kb	7.11 (0.38)
DNAmTL residuals	0 (0.38)
Skin and Blood Clock Epigenetic age (years)	15.7 (1.94)
Skin and blood clock residuals	0 (1.77)

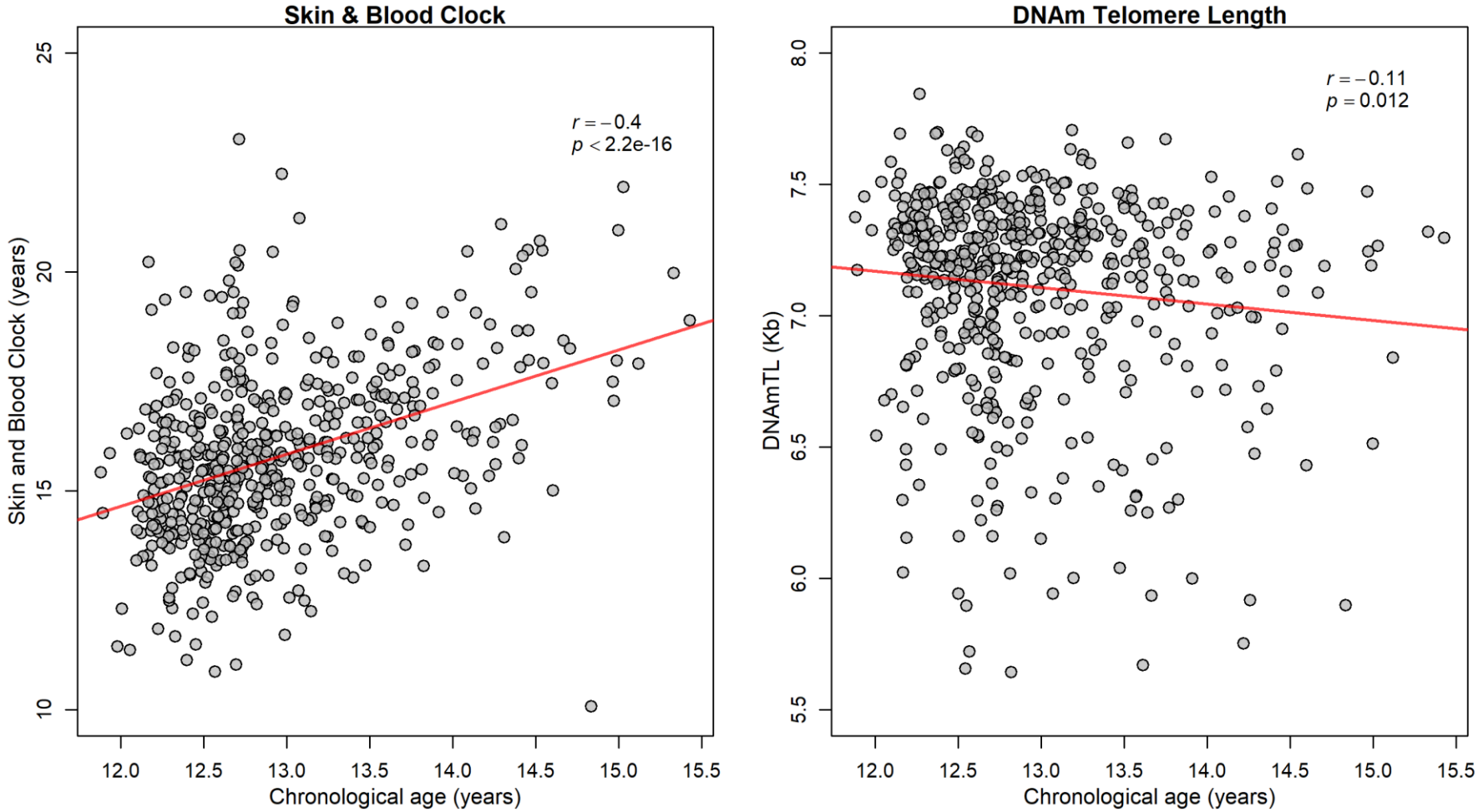
Table S2. Differences in %-DNAm of the *ACE2* gene in nasal cells by chronological age in years stratified by sex

CpG	Females				Males		
	%-DNAm difference	*p*	FDR		%-DNAm difference	*p*	FDR
cg18458833	-0.02	0.76	0.89		-0.11	0.43	0.89
cg21598868	-0.89	0.15	0.32		-0.04	0.77	0.89
cg18877734	-0.34	0.06	0.18		-0.18	0.34	0.86
cg03536816	**-2.48**	**0.018**	0.16		-0.91	0.29	0.86
cg08559914	0.01	0.96	0.96		-0.24	0.49	0.89
cg16734967	**-1.85**	**0.048**	0.18		-0.62	0.29	0.86
cg10408040	-0.1	0.66	0.83		0.03	0.73	0.89
cg05241917	-0.42	0.19	0.36		-0.12	0.62	0.89
cg20119767	-0.21	0.59	0.81		0.34	0.31	0.86
cg04013915	-1.48	0.07	0.18		0.05	0.93	0.93
cg16716680	-0.03	0.95	0.96		0.17	0.75	0.89
cg25176872	-0.56	0.23	0.39		0.01	0.93	0.93
cg22422976	**-1.19**	**0.02**	0.16		-0.04	0.75	0.89
cg05039749	-0.11	0.34	0.52		-0.05	0.19	0.86
cg23232263	-0.96	0.04	0.18		-0.30	0.05	0.80

Table S3. Differences in %-DNAm in Asian and more than one race relative to white participants adjusted for age and stratified by sex.

	Females								**Males**						
CpG	**Asian**				**More than one race**				**Asian**				**More than one race**		
	%-DNAm difference	*p*	FDR		%-DNAm difference	*p*	FDR		%-DNAm difference	*p*	FDR		%-DNAm difference	*p*	FDR
cg18458833	-0.22	0.27	0.87		-0.14	0.41	0.56		0.33	0.48	0.96		-0.35	0.45	0.79
cg21598868	-0.36	0.87	0.87		1.72	0.25	0.56		0.20	0.66	0.96		-0.27	0.47	0.79
cg18877734	-0.28	0.66	0.87		0.1	0.79	0.88		0.10	0.89	0.96		-0.3	0.58	0.79
cg03536816	-1.54	0.72	0.87		-4.8	0.17	0.56		4.30	0.03	0.38		-0.44	0.79	0.79
cg08559914	0.12	0.85	0.87		-0.4	0.33	0.56		0.58	0.49	0.96		-0.76	0.33	0.79
cg16734967	1.81	0.55	0.87		-3.05	0.19	0.56		0.87	0.65	0.96		-0.49	0.65	0.79
cg10408040	0.87	0.33	0.87		-0.5	0.40	0.56		-0.47	0.31	0.96		-0.38	0.10	0.79
cg05241917	0.57	0.62	0.87		0.17	0.82	0.88		0.33	0.64	0.96		0.18	0.69	0.79
cg20119767	1.53	0.17	0.87		1.23	0.33	0.56		-0.05	0.96	0.96		-0.24	0.75	0.79
cg04013915	1.89	0.59	0.87		-2.55	0.35	0.56		0.75	0.55	0.96		-0.46	0.68	0.79
cg16716680	-0.32	0.83	0.87		1.29	0.45	0.56		-0.13	0.95	0.96		0.75	0.32	0.79
cg25176872	1.03	0.16	0.87		-1.22	0.21	0.56		0.02	0.96	0.96		-0.09	0.74	0.79
cg22422976	-0.4	0.79	0.87		0.01	0.99	0.99		0.31	0.42	0.96		-0.23	0.41	0.79
cg05039749	0.47	0.50	0.87		-0.24	0.25	0.56		0.01	0.93	0.96		0.11	0.15	0.79
cg23232263	-0.39	0.82	0.87		-0.84	0.43	0.56		0.16	0.64	0.96		-0.42	0.19	0.79

Figure S1. Scatter plots, and Pearson's correlation coefficients between children's chronological age and DNA methylation aging biomarkers A) Skin & Blood Clock and B) DNA methylation estimator of telomere length (DNAmTL) in nasal cells.



References

1 Hou, Y. J. *et al.* SARS-CoV-2 Reverse Genetics Reveals a Variable Infection Gradient in the Respiratory Tract. *Cell* (2020).

2 Sungnak, W. *et al.* SARS-CoV-2 entry factors are highly expressed in nasal epithelial cells together with innate immune genes. *Nature medicine* **26**, 681-687 (2020).

3 World Health Organization. *Coronavirus disease (COVID-19) pandemic*, <https://www.who.int/emergencies/diseases/novel-coronavirus-2019> (2020).

4 Van Doremalen, N. *et al.* Aerosol and surface stability of SARS-CoV-2 as compared with SARS-CoV-1. *New England Journal of Medicine* **382**, 1564-1567 (2020).

5 Klompas, M., Baker, M. A. & Rhee, C. Airborne transmission of SARS-CoV-2: theoretical considerations and available evidence. *JAMA* (2020).

6 Hoffmann, M. *et al.* SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell* (2020).

7 Zhou, P. *et al.* A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270-273 (2020).

8 Smith, J. C. *et al.* Cigarette smoke exposure and inflammatory signaling increase the expression of the SARS-CoV-2 receptor ACE2 in the respiratory tract. *Developmental Cell* (2020).

9 McCray, P. B. *et al.* Lethal infection of K18-hACE2 mice infected with severe acute respiratory syndrome coronavirus. *Journal of virology* **81**, 813-821 (2007).

10 Benmarhnia, T. Linkages between air pollution and the health burden from COVID-19: methodological challenges and opportunities. *American journal of epidemiology* (2020).

11 Wu, X., Nethery, R. C., Sabath, B. M., Braun, D. & Dominici, F. Exposure to air pollution and COVID-19 mortality in the United States. *medRxiv* (2020).

12 Lin, C.-I. *et al.* Instillation of particulate matter 2.5 induced acute lung injury and attenuated the injury recovery in ACE2 knockout mice. *International journal of biological sciences* **14**, 253 (2018).

13 Aztatzi-Aguilar, O. G., Uribe-Ramírez, M., Arias-Montaño, J. A., Barbier, O. & De Vizcaya-Ruiz, A. Acute and subchronic exposure to air particulate matter induces expression of angiotensin and bradykinin-related genes in the lungs and heart: Angiotensin-II type-I receptor as a molecular target of particulate matter exposure. *Particle and fibre toxicology* **12**, 17 (2015).

14 Leung, J. M. *et al.* ACE-2 expression in the small airway epithelia of smokers and COPD patients: implications for COVID-19. *European Respiratory Journal* **55** (2020).

15 Sawalha, A. H., Zhao, M., Coit, P. & Lu, Q. Epigenetic dysregulation of ACE2 and interferon-regulated genes might suggest increased COVID-19 susceptibility and severity in lupus patients. *Clinical Immunology*, 108410 (2020).

16 Zill, P. *et al.* DNA methylation analysis of the angiotensin converting enzyme (ACE) gene in major depression. *PLoS One* **7**, e40479 (2012).

17 Chlamydas, S., Papavassiliou, A. G. & Piperi, C. Epigenetic mechanisms regulating COVID-19 infection. *Epigenetics* (2020).

18 Pruimboom, L. Methylation Pathways and SARS-CoV-2 Lung Infiltration and Cell Membrane-Virus Fusion Are Both Subject to Epigenetics. *Frontiers in Cellular and Infection Microbiology* **10**, 290 (2020).

19 Tukiainen, T. *et al.* Landscape of X chromosome inactivation across human tissues. *Nature* **550**, 244-248 (2017).

20 Liu, J. *et al.* Sex differences in renal angiotensin converting enzyme 2 (ACE2) activity are 17β-oestradiol-dependent and sex chromosome-independent. *Biology of sex differences* **1**, 6 (2010).

21 Bezara, M. E. O. *et al.* Heterogeneous expression of the SARS-Coronavirus-2 receptor ACE2 in the human respiratory tract. *bioRxiv* (2020).

22 Zou, L. *et al.* SARS-CoV-2 viral load in upper respiratory specimens of infected patients. *New England Journal of Medicine* **382**, 1177-1179 (2020).

23 Chen, M. *et al.* Elevated ACE2 expression in the olfactory neuroepithelium: implications for anosmia and upper respiratory SARS-CoV-2 entry and replication. *bioRxiv* (2020).
24 Pinto, B. G. *et al.* ACE2 expression is increased in the lungs of patients with comorbidities associated with severe COVID-19. *MedRxiv* (2020).
25 Zhao, Y. *et al.* Single-cell RNA expression profiling of ACE2, the putative receptor of Wuhan 2019-nCov. *BioRxiv* (2020).
26 Chen, Y., Shan, K. & Qian, W. Asians do not exhibit elevated expression or unique genetic polymorphisms for ACE2, the cell-entry receptor of SARS-CoV-2. *Preprint at doi* **10** (2020).
27 Corley, M. J. & Ndhlovu, L. C. DNA methylation analysis of the COVID-19 host cell receptor, angiotensin I converting enzyme 2 gene (ACE2) in the respiratory system reveal age and gender differences. (2020).
28 Bunyavanich, S., Do, A. & Vicencio, A. Nasal gene expression of angiotensin-converting enzyme 2 in children and adults. *Jama* (2020).
29 Kim L, W. M., O'Halloran A, et al. . Hospitalization Rates and Characteristics of Children Aged <18 Years Hospitalized with Laboratory-Confirmed COVID-19 — COVID-NET, 14 States, March 1–July 25, 2020. *MMWR Morb Mortal Wkly Rep* **ePub: 7 August 2020.**, doi:http://dx.doi.org/10.15585/mmwr.mm6932e3external icon (2020).
30 Cao, Y. *et al.* Comparative genetic analysis of the novel coronavirus (2019-nCoV/SARS-CoV-2) receptor ACE2 in different populations. *Cell discovery* **6**, 1-4 (2020).
31 Stawiski, E. W. *et al.* Human ACE2 receptor polymorphisms predict SARS-CoV-2 susceptibility. *BioRxiv*, doi:https://doi.org/10.1101/2020.04.07.024752 (2020).
32 Oken, E. *et al.* Cohort profile: project viva. *International journal of epidemiology* **44**, 37-48 (2015).
33 Lai, P. S. *et al.* Alternate methods of nasal epithelial cell sampling for airway genomic studies. *Journal of Allergy and Clinical Immunology* **136**, 1120-1123. e1124 (2015).
34 Cardenas, A. *et al.* The nasal methylome as a biomarker of asthma and airway inflammation in children. *Nature communications* **10**, 1-10 (2019).
35 Fortin, J.-P. *et al.* Functional normalization of 450k methylation array data improves replication in large cancer studies. *Genome biology* **15**, 503 (2014).
36 Niu, L., Xu, Z. & Taylor, J. A. RCP: a novel probe design bias correction method for Illumina Methylation BeadChip. *Bioinformatics* **32**, 2659-2663 (2016).
37 Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882-883 (2012).
38 Lu, A. T. *et al.* DNA methylation-based estimator of telomere length. *Aging (Albany NY)* **11**, 5895 (2019).
39 Horvath, S. *et al.* Epigenetic clock for skin and blood cells applied to Hutchinson Gilford Progeria Syndrome and ex vivo studies. *Aging (Albany NY)* **10**, 1758 (2018).
40 Martin, T. C., Yet, I., Tsai, P.-C. & Bell, J. T. coMET: visualisation of regional epigenome-wide association scan results and DNA co-methylation patterns. *BMC bioinformatics* **16**, 1-5 (2015).