

UC Berkeley

UC Berkeley Previously Published Works

Title

DNA methylation architecture of the ACE2 gene in nasal cells of children

Permalink

<https://escholarship.org/uc/item/3z3948q7>

Journal

Scientific Reports, 11(1)

ISSN

2045-2322

Authors

Cardenas, Andres

Rifas-Shiman, Sheryl L

Sordillo, Joanne E

et al.

Publication Date

2021

DOI

10.1038/s41598-021-86494-7

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed



OPEN

DNA methylation architecture of the *ACE2* gene in nasal cells of children

Andres Cardenas^{1,2,✉}, Sheryl L. Rifas-Shiman³, Joanne E. Sordillo³, Dawn L. DeMeo⁴, Andrea A. Baccarelli⁵, Marie-France Hivert^{3,6}, Diane R. Gold^{4,7,8} & Emily Oken^{3,8}

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} The nasal epigenome is in direct contact with the environment and could explain COVID-19 disparities by reflecting social and environmental influences on *ACE2* regulation. 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 relative to males. We observed differential DNAm at 5 CpGs for Hispanic females (mean absolute difference = 3.22%) and lower DNAm at 8 CpGs for Black males (mean absolute difference = 1.33%), relative to white participants. Longer DNAm telomere length was associated with greater *ACE2* 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 reflecting upstream environmental and social influences. Findings need to be confirmed among adults and patients with risk factors for COVID-19 severity.

The global COVID-19 pandemic caused by SARS-CoV-2 has infected over 120 million individuals and claimed over 2,674,000 lives worldwide as of March 18, 2021³. 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 severity⁸. In a different strain of SARS coronavirus, transgenic mice expressing human ACE2 in airway and other epithelia cells develop a rapidly lethal infection⁹.

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 lungs¹. 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 females¹⁸ with heterogeneous but significant male-bias in expression in several tissues¹⁹. However, no study has examined the DNAm

¹Division of Environmental Health Sciences, School of Public Health, University of California, Berkeley, 2121 Berkeley Way, #5121, Berkeley, CA 94720, USA. ²Center for Computational Biology, University of California, Berkeley, Berkeley, CA, USA. ³Division of Chronic Disease Research Across the Lifecourse, Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, MA, USA. ⁴Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. ⁵Department of Environmental Health Sciences, Mailman School of Public Health, Columbia University, New York, NY, USA. ⁶Diabetes Unit, Massachusetts General Hospital, Boston, MA, USA. ⁷Department of Environmental Health, Harvard T.H. Chan School of Public Health, Boston, MA, USA. ⁸These authors contributed equally: Diane R. Gold and Emily Oken. ✉email: andres.cardenas@berkeley.edu

CpG Name	Position	Gene Group	TFBS coordinates X-chromosomes	TFBS evidence count	Mean (SD) %-DNA methylation			†Adjusted %-DNAm difference	‡Yang et al 450K ²⁰
					Overall	Females	Males		
cg18458833	15,621,477	TSS1500			95.47 (1.84)	96.28 (0.70)	94.68 (1.84)	1.52	3.44
cg21598868	15,621,167	TSS1500			87.88 (2.51)	81.56 (6.08)	94.05 (2.51)	-12.12	-4.32
cg18877734	15,621,084	TSS1500			94.26 (2.22)	94.00 (2.14)	94.51 (2.22)	-0.45	3.09
cg03536816	15,621,042	TSS1500;Body			66.55 (12.88)	56.14 (12.14)	76.69 (12.88)	-21.29	
cg08559914	15,620,240	TSS200			91.80 (4.84)	92.98 (2.88)	90.65 (4.84)	1.84	7.89
cg16734967	15,620,103	1stExon;5'UTR	15,619,908–15,620,856	6	77.51 (8.28)	72.46 (9.57)	82.44 (8.28)	-9.76	
cg10408040	15,619,136	5'UTR;ExonBnd			92.76 (1.30)	91.05 (2.75)	94.42 (1.30)	-3.05	
cg05241917	15,616,197	Body	15,616,057–15,616,589	4	91.16 (4.48)	89.78 (4.90)	92.49 (4.48)	-2.31	
cg20119767	15,612,969	ExonBnd;Body	15,612,415–15,612,981	6	83.72 (5.01)	79.52 (4.90)	87.82 (5.01)	-8.67	-5.55
cg04013915	15,586,143	Body	15,586,103–15,586,918	6	67.24 (10.23)	51.64 (10.36)	82.44 (10.23)	-31.47	
cg16716680	15,585,368	Body			76.87 (5.76)	66.82 (6.54)	86.68 (5.76)	-20.08	
cg25176872	15,584,988	Body	15,582,921–15,585,025	6	89.64 (2.41)	85.20 (4.37)	93.97 (2.41)	-8.49	
cg22422976	15,584,867	Body	15,582,921–15,585,025	6	81.54 (2.42)	68.72 (4.97)	94.04 (4.42)	-25.22	
cg05039749	15,583,512	Body			2.53 (0.45)	3.16 (1.47)	1.91 (0.45)	1.00	-1.29
cg23232263	15,579,482	3'UTR			90.23 (3.72)	85.52 (5.73)	94.81 (3.72)	-8.96	-7.01

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

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. No differences in chronological age were observed by sex ($p > 0.05$). Black participants were marginally older ($\beta = 0.13$ years; 95% CI: -0.01, 0.29; $p = 0.08$) compared to white participants.

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 to 6 experiments extracted from Encyclopedia of DNA Elements (ENCODE) supporting transcription binding at that region. After excluding one CpG due to low detection performance, DNAm levels of the other 15 CpGs ranged from 2.53% to 95.47% and 14 sites had DNAm levels above 60%, all with robust significant differences by sex (FDR < 0.003), Table 1 and Fig. 1A. 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 -0.45% at cg18877734, Table 1. 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, Fig. 1B. 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. Using an external data of 72 children aged 10–12 years of nasal DNA measured with the 450 K Illumina array by Yang and colleagues²⁰ we observed consistent strong differences by sex at cg18458833; cg21598868; cg08559914; cg20119767 and cg23232263. However, two sites had significant differences by sex (cg18877734 and cg05039749) in the opposite direction compared to our data. Of note is that these two CpGs showed the weakest differences by sex in our data, Table 1. This cohort was predominately Black not allowing us to test differences by race/ethnicity and half of the children had asthma which could have influenced the results at the two sites.

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 Fig. 1C. 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

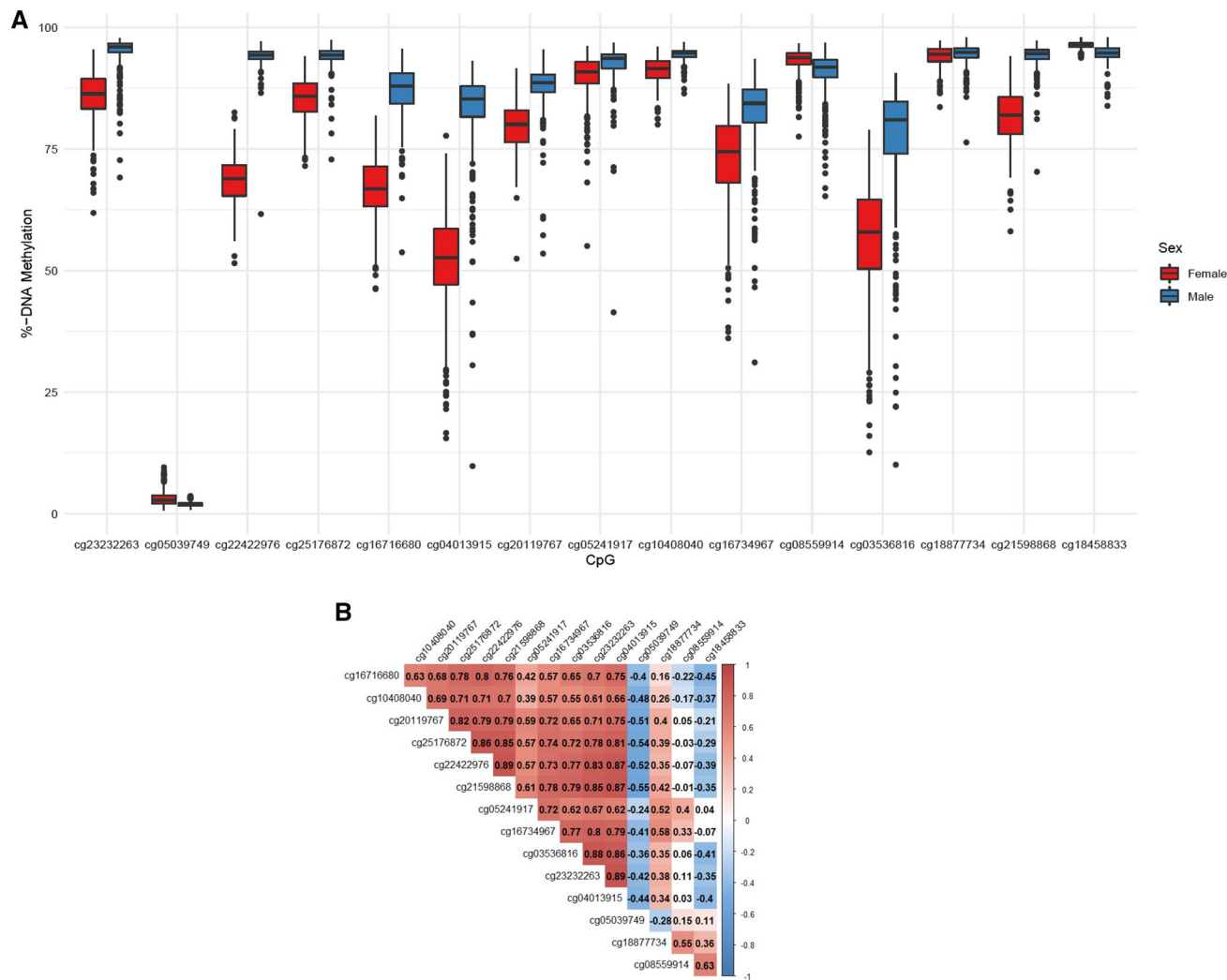


Figure 1. (A) Boxplots of DNA methylation in nasal cells for the angiotensin-converting enzyme 2 (*ACE2*) gene by sex among 15 CpG sites. (B) Spearman correlation coefficients of nasal cell DNA methylation levels among CpGs found in the angiotensin-converting enzyme (*ACE2*) gene. (C) Local Manhattan plot for the association of *ACE2* DNA methylation for Hispanic relative to white females. Spearman correlation matrix for CpGs among all females. (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.

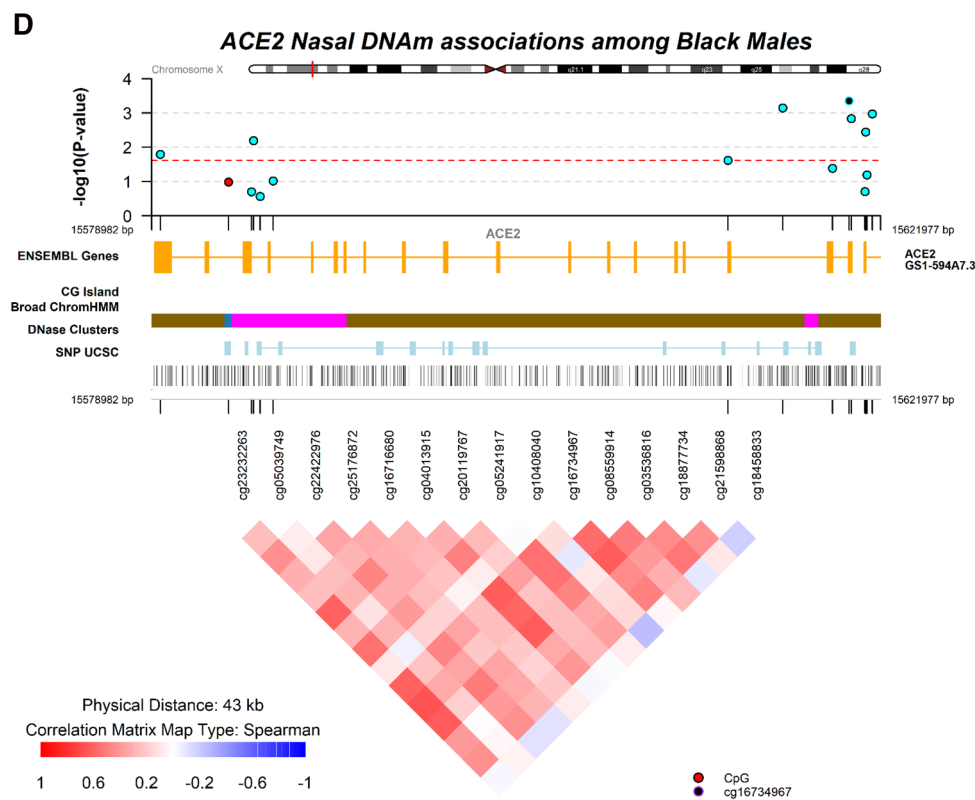
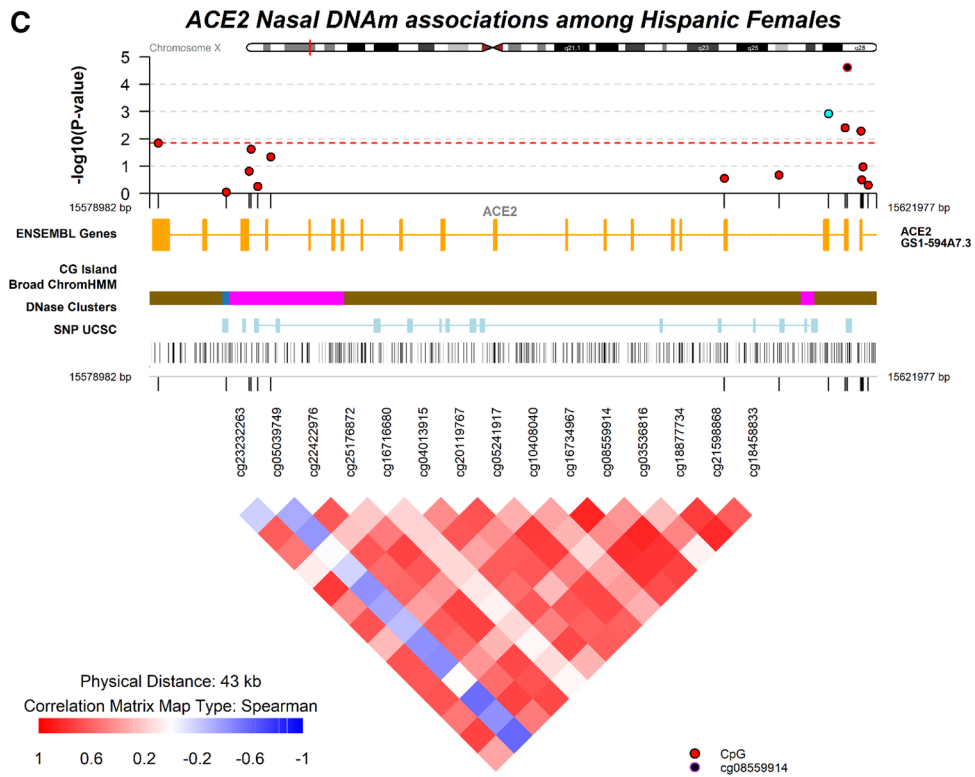


Figure 1. (continued)

CpG	Females						Males					
	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 2. Differences in %-DNAm in Black and Hispanic relative to white participants adjusted for age and stratified by sex. Bold values indicate FDR <0.05.

CpG	DNAmTLadjAge						Skin & Blood Clock Acceleration					
	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	1.24 × 10⁻¹¹	2.32 × 10⁻¹¹	3.17	6.13 × 10⁻¹⁹	1.31 × 10⁻¹⁸	0.39	0.05	0.08	-0.01	0.92	0.92
cg18877734	2.13	1.06 × 10⁻¹⁰	1.77 × 10⁻¹⁰	2.48	4.43 × 10⁻⁸	7.39 × 10⁻⁸	0.15	0.05	0.08	0.05	0.51	0.92
cg03536816	24.34	5.06 × 10⁻¹²⁴	7.59 × 10⁻¹²³	28.2	6.50 × 10⁻⁵³	9.74 × 10⁻⁵²	0.97	0.02	0.05	0.15	0.72	0.92
cg08559914	4.44	2.14 × 10⁻²⁵	6.43 × 10⁻²⁵	7.41	8.60 × 10⁻²⁷	4.30 × 10⁻²⁶	0.19	0.03	0.06	-0.03	0.83	0.92
cg16734967	17.32	2.87 × 10⁻⁵⁰	1.08 × 10⁻⁴⁹	11.68	1.59 × 10⁻¹⁹	3.97 × 10⁻¹⁹	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	9.62 × 10⁻¹⁵	2.40 × 10⁻¹⁴	4.18	1.77 × 10⁻²⁸	1.33 × 10⁻²⁷	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	2.31 × 10⁻⁷²	1.16 × 10⁻⁷¹	18.2	7.48 × 10⁻²⁰	2.24 × 10⁻¹⁹	1.15	0.004	0.03	0.04	0.88	0.92
cg16716680	4.15	5.52 × 10⁻⁶	7.52 × 10⁻⁶	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	9.18 × 10⁻⁵	0.00014	0.22	0.16	0.18	-0.01	0.86	0.92
cg22422976	5.03	3.67 × 10⁻¹³	7.86 × 10⁻¹³	1.55	9.13 × 10⁻¹³	1.71 × 10⁻¹²	0.18	0.36	0.38	-0.08	0.15	0.80
cg05039749	0.72	7.00 × 10⁻⁷	1.05 × 10⁻⁶	-0.08	0.25	0.27	0.09	0.03	0.06	0.01	0.57	0.92
cg23232263	11.04	4.50E × 10⁻⁸⁸	3.37 × 10⁻⁸⁷	6.05	7.67 × 10⁻²⁵	2.87 × 10⁻²⁴	0.37	0.08	0.11	0.06	0.38	0.92

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. Bold values indicate FDR <0.05. *Adjusted for race/ethnicity.

annotation and correlation structure is presented in Fig. 1D. No differences were observed among Asian or participants of more than one race, Table S3.

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 with significant correlations of $r = -0.11$ and $r = 0.40$, 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 for females, Table 3.

Hierarchical clustering of the DNAm of ACE2 CpGs confirmed the large influence of sex across all DNAm levels and a relatively weak influence of age and race/ethnicity, Figure S2.

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. Our findings need to be confirmed among adults with a wider age range and associations should be tested for known risk factors of COVID-19 severity.

The *ACE2* gene is located in the X-chromosome and escapes X-inactivation in females with potential tissue specificity and heterogeneous male expression bias, partially attributed to sex steroids²¹. 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^{2,22}. 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^{1,23}. Increased *ACE2* expression has been hypothesized to partially explain differences in COVID-19 severity²⁴. 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* expression²⁵.

ACE2 expression has been reported to be higher in Asians compared with white and African Americans from type II alveolar cells²⁶. However, other reports have found no difference in *ACE2* expression by race²⁷. 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 cells²⁸. 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 in our sample. However, expression of *ACE2* in the lung has been demonstrated to increase with age²⁷. In nasal samples, lower *ACE2* expression among children relative to older adults has been reported²⁹. 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 children³⁰. 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^{31,32}. Therefore, we hypothesize that observed DNA methylation variation originates from social and environmental exposures which include social and environmental racism.

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 participant 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. Young children have lower rates of COVID-19 infection compared to adults^{33,34}. Therefore, the generalizability among adults and implication related to COVID-19 severity need to be further investigated. Study designs that incorporate functional genomics with known risk factors such as co-morbid conditions among adults and patients with varying degrees of COVID-19 severity could yield important insight among this population for which the risk of severe disease is greater³⁴. Additionally the systematic investigation beyond nasal tissue in leukocytes across diverse cohorts could yield insights into severity and infection risk. 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. Additionally, while the magnitude of DNA methylation differences by sex were moderate, smaller differences were observed by race/ethnicity. Small to moderate differences in DNA methylation have been shown to be functionally relevant³⁵. Even if absolute changes are small, they must be interpreted relative to mean levels. The accuracy of the array might vary by probe and experiment. However, the EPIC array has been demonstrated to be highly reproducible across technical and biological replicates across probes³⁶.

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. These findings need to be confirmed in adults and patients with risk factors for severe COVID-19.

Materials and 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 Associates³⁷. 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 post-

partum follow-up visits. Written informed consent was obtained from parent and/or legal guardian of minors and children provided written assent. Institutional Review Board of Harvard Pilgrim Health Care reviewed and approved all study protocols. All methods were carried out in accordance with relevant guidelines and regulations stipulated by the institutional review board.

Nasal DNA methylation analyses. Trained research staff collected nasal swabs from the anterior nares, previously shown to yield respiratory epithelial cells³⁸. 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 described³⁹. We preprocessed all of the DNA methylation data using functional normalization to adjust for technical variability⁴⁰ and adjusted for probe-type variability using RCP⁴¹. Lastly, we used ComBat from the *sva* package to adjust for sample plate (seven 96-well plates) as a technical variable⁴². 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 ($p < 1 \times 10^{-8}$) 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)⁴³ and the "skin and blood epigenetic clock"⁴⁴. 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.

External replication. We identified an external dataset in the Gene Expression Omnibus (GEO) repository (GSE65205). The data contained 72 samples from mostly African American children (>90%) with persistent atopic asthma ($n = 36$) versus healthy controls ($n = 36$) aged 10 to 12 years with half of the samples from each sex²⁰. We extracted data from GEO and preprocessed DNA methylation data in the same fashion with the exception that this study used the older Illumina methylation 450 K array and therefore only 7 CpGs overlapped with the EPIC measurements. We tested associations by sex with the same approach as our data.

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 coMET⁴⁵. 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/).

Data availability

Datasets generated and analyzed during the current study are not publicly available because we did not obtain consent for such public release of epigenetic data from participants. Raw data to generate figures and tables are available from the corresponding author with the appropriate permission from the Project Viva study team and investigators (project_viva@hphc.org) upon reasonable request and institutional review board approval.

Received: 9 September 2020; Accepted: 17 March 2021

Published online: 29 March 2021

References

- Hou, Y. J. *et al.* SARS-CoV-2 reverse genetics reveals a variable infection gradient in the respiratory tract. *Cell* **182**, 429–446. e414 (2020).
- Sungnak, W. *et al.* SARS-CoV-2 entry factors are highly expressed in nasal epithelial cells together with innate immune genes. *Nat. Med.* **26**, 681–687 (2020).
- World Health Organization. *Coronavirus disease (COVID-19) pandemic*, <https://www.who.int/emergencies/diseases/novel-coronavirus-2019> (2021).
- Van Doremalen, N. *et al.* Aerosol and surface stability of SARS-CoV-2 as compared with SARS-CoV-1. *N. Engl. J. Med.* **382**, 1564–1567 (2020).

5. Klompas, M., Baker, M. A. & Rhee, C. Airborne transmission of SARS-CoV-2: theoretical considerations and available evidence. *JAMA* **324**, 441–442 (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* **181**, 271–280. e278 (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* **53**, 514–529. e513 (2020).
9. McCray, P. B. *et al.* Lethal infection of K18-hACE2 mice infected with severe acute respiratory syndrome coronavirus. *J. Virol.* **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* **189**, 1238–1243 (2020).
11. Wu, X., Nethery, R. C., Sabath, M., Braun, D. & Dominici, F. Air pollution and COVID-19 mortality in the United States: Strengths and limitations of an ecological regression analysis. *Science Advances* **6**, eabd4049 (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-Montaña, 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. *Part. Fibre Toxicol.* **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**, 2000688. <https://doi.org/10.1183/13993003.00688-2020> (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* **16**, 263–270 (2021).
18. Pruumboom, L. Methylation Pathways and SARS-CoV-2 Lung Infiltration and Cell Membrane-Virus Fusion Are Both Subject to Epigenetics. *Front. Cell. Infect. Microbiol.* **10**, 290 (2020).
19. Tukiainen, T. *et al.* Landscape of X chromosome inactivation across human tissues. *Nature* **550**, 244–248 (2017).
20. Yang, I. V. *et al.* The nasal methylome and childhood atopic asthma. *J. Allergy Clin. Immunol.* **139**, 1478–1488 (2017).
21. Liu, J. *et al.* Sex differences in renal angiotensin converting enzyme 2 (ACE2) activity are 17 β -oestradiol-dependent and sex chromosome-independent. *Biol. Sex Differ.* **1**, 6 (2010).
22. Ortiz, M. E. *et al.* Heterogeneous expression of the SARS-Coronavirus-2 receptor ACE2 in the human respiratory tract. *EBioMedicine* **60**, 102976 (2020).
23. Zou, L. *et al.* SARS-CoV-2 viral load in upper respiratory specimens of infected patients. *N. Engl. J. Med.* **382**, 1177–1179 (2020).
24. Chen, M. *et al.* Elevated ACE-2 expression in the olfactory neuroepithelium: implications for anosmia and upper respiratory SARS-CoV-2 entry and replication. *Eur. Respir. J.* **56** (2020).
25. Pinto, B. G. *et al.* ACE2 expression is increased in the lungs of patients with comorbidities associated with severe COVID-19. *J. Infect. Dis.* **222**, 556–563 (2020).
26. Zhao, Y. *et al.* Single-cell RNA expression profiling of ACE2, the receptor of SARS-CoV-2. *Am. J. Respir. Crit. Care Med.* **202**, 756–759 (2020).
27. 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. *Preprints 2020*, 2020020258. <https://doi.org/10.20944/preprints202002.0258.v2>.
28. 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. *Preprints 2020*, 2020030295. <https://doi.org/10.20944/preprints202003.0295.v1>.
29. Bunyavanich, S., Do, A. & Vicencio, A. Nasal gene expression of angiotensin-converting enzyme 2 in children and adults. *JAMA* **323**, 2427–2429 (2020).
30. 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**. <https://doi.org/10.15585/mmwr.mm6932e3external>.
31. 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).
32. Stawiski, E. W. *et al.* Human ACE2 receptor polymorphisms predict SARS-CoV-2 susceptibility. *BioRxiv* <https://doi.org/10.1101/2020.04.07.024752> (2020).
33. Viner, R. M. *et al.* Susceptibility to SARS-CoV-2 infection among children and adolescents compared with adults: a systematic review and meta-analysis. *JAMA pediatrics* (2020).
34. Goldstein, E., Lipsitch, M. & Cevik, M. On the Effect of Age on the Transmission of SARS-CoV-2 in Households, Schools, and the Community. *J. Infect. Dis.* **223**, 362–369 (2021).
35. Breton, C. V. *et al.* Small-magnitude effect sizes in epigenetic end points are important in children's environmental health studies: the children's environmental health and disease prevention research center's epigenetics working group. *Environ. Health Perspect.* **125**, 511–526 (2017).
36. Pidsley, R. *et al.* Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. *Genome Biol.* **17**, 1–17 (2016).
37. Oken, E. *et al.* Cohort profile: project viva. *Int. J. Epidemiol.* **44**, 37–48 (2015).
38. Lai, P. S. *et al.* Alternate methods of nasal epithelial cell sampling for airway genomic studies. *J. Allergy Clin. Immunol.* **136**, 1120–1124 (2015).
39. Cardenas, A. *et al.* The nasal methylome as a biomarker of asthma and airway inflammation in children. *Nat. Commun.* **10**, 1–10 (2019).
40. Fortin, J.-P. *et al.* Functional normalization of 450k methylation array data improves replication in large cancer studies. *Genome Biol.* **15**, 503 (2014).
41. Niu, L., Xu, Z. & Taylor, J. A. RCP: a novel probe design bias correction method for Illumina Methylation BeadChip. *Bioinformatics* **32**, 2659–2663 (2016).
42. 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).
43. Lu, A. T. *et al.* DNA methylation-based estimator of telomere length. *Aging (Albany NY)* **11**, 5895 (2019).
44. 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).
45. 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).

Author contributions

A.C. performed the statistical analyses and wrote the main manuscript. D.R.G. and E.O. conceived the initial study and provided critical feedback on the manuscript. S.L.R.S., J.E.S., D.L.D., A.A.B. and M.F.H. provided critical input on analyses and interpretation of the data. All authors reviewed the manuscript.

Funding

This work was supported by the US National Institutes of Health Grants R01 ES031259, R01 HD034568 and UH3OD023286.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-86494-7>.

Correspondence and requests for materials should be addressed to A.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021