Epigenome-wide association studies of three social determinants of health and implications for lung functions among survivors of childhood cancer

Nan Song^{1,*}, Jin-ah Sim^{1,*}, Qian Dong¹, Yinan Zheng², Lifang Hou², Zhenghong Li¹, Chia-Wei Hsu³, Haitao Pan³, Heather Mulder⁴, John Easton⁴, Emily Walker⁵, Geoffrey Neale⁵, Carmen L. Wilson¹, Kirsten K. Ness¹, Kevin R. Krull¹, Deo Kumar Srivastava³, Yutaka Yasui¹, Jinghui Zhang⁴, Melissa M. Hudson^{1,6}, Leslie L. Robison¹, I-Chan Huang^{1,#}, Zhaoming Wang^{1,4,#}

¹*Department of Epidemiology and Cancer Control, St. Jude Children's Research Hospital, Memphis, TN, USA;* ²*Department of Preventive Medicine, Northwestern University, Chicago, IL, USA;* ³*Department of Biostatistics,* ⁴*Department of Computational Biology,* ⁵*Hartwell Center,* ⁶*Department of Oncology, St. Jude Children's Research Hospital, Memphis, TN, USA*

*N.S. and J.S. contributed equally as co-first authors.

#Z.W. and I-C.H. contributed equally as co-senior authors.

Correspondence: Zhaoming Wang, Department of Epidemiology and Cancer Control, St. Jude Children's Research Hospital, 262 Danny Thomas Place, MS 735, Memphis, TN 38105 USA, Phone: 901-595-0232, Email: Zhaoming.Wang@stjude.org

Abstract

Background: Emerging evidence suggests that social determinants of health (SDOH) may influence health and wellness through an epigenetic mechanism in the general population. However, the social epigenomic approach has not yet been applied to survivors of childhood cancer, a vulnerable population with elevated risk for chronic health conditions (CHCs).

Methods: Study participants were drawn from the St Jude Lifetime Cohort, a hospital-based retrospective cohort with prospective follow up. DNA methylation (DNAm) profiling was generated based on blood derived DNA collected during follow-up visit. SDOH included educational attainment, personal income, and area deprivation index (ADI) based on baseline or follow-up questionnaires and geocoding. CHCs were clinically assessed with severity grade.

Results: We included 258 childhood cancer survivors of African ancestry (AA) (median time from diagnosis=25.2 years, interquartile range [IQR]=19.9-32.1 years) and 1,618 survivors of European ancestry (EA) (median time from diagnosis=27.3, IQR=21.1-33.7 years). Through epigenome-wide association studies, we identified 130 SDOH-CpG associations including educational attainment (N=88), personal income (N=23), and ADI (N=19) at epigenome-wide significance level ($P < 9 \times 10^{-8}$). There were 13 CpGs, commonly associated with all three SDOH factors, with attenuated remaining effect sizes (36.8-48.3%) after additionally adjusting body mass index and smoking, mapped to smoking-related genes including *GPR55*, *CLDND1*, *CPOX*, *GPR15*, *AHRR*, *PRRC2B*, and *ELMSAN1*. Among 130 SDOH-related CpGs, three independent CpGs (cg04180924, cg1120500, and cg27470486) had a significant combined mediation effect for educational attainment (%mediation=48.9%), and a single mediator cg08064403 was found with significant mediation effect for personal income (25.9%) and ADI (24.1%) on pulmonary

diffusion deficit, which showed higher incidence in AA than in EA survivors implying racial disparity which is possibly due to more disadvantageous SDOH factors in AA than in EA.

Conclusions: We demonstrated striking DNAm signatures associated with multiple SDOH factors (educational attainment, personal income, and ADI) and many epigenome-wide significant CpG sites resembling the effect of smoking exposure. We also identified an exemplified racial health disparity in pulmonary diffusion deficit between AA and EA survivors and illuminated DNAm as potential mechanistic mediators for SDOH factors using a social epigenomic approach.

Introduction

Previous research suggested that childhood cancer survivors of African ancestry (AA) had significantly poorer survivorship including higher mortality and prevalent cardiovascular disorders, stroke, diabetes, hypertension, and neurocognitive and emotional outcomes compared to non-Hispanic Caucasian survivors.¹⁻⁴ However, after adjusting for the influence of personal socio-economic factors, health disparities were either substantially decreased or became not statistically significant,¹⁻⁴ suggesting the importance of socio-economic factors in racial health disparity among survivors of childhood cancer.

Social and behavioral epigenomics is an emerging transdisciplinary field studying social and the relevant health behavioral factors that can influence health outcomes through an epigenetic mechanism and how they may contribute to health disparities.^{5,6} Among a variety of epigenetic modifications, DNA methylation (DNAm) has been the most widely studied, reporting significant associations with overall socio-economic status (SES),^{5,7-11} educational attainment,^{12,13} and health behaviors.¹⁴ The epigenetic signatures may further link SES to chronic disease risk.¹⁰ Recently, we reported the large-scale epigenetic study, showing that cancer treatment exposures were associated with persistent variations of methylation in DNA derived from blood decades later and a subset of treatment-associated DNAm 5'-cytosine-phosphate-guanine-3' (CpG) sites was also significantly associated with cardiometabolic risk.¹⁵ Although intensive treatment exposures would be expected to leave persistent epigenetic marks and shape up the survivors' methylome, poor social and behavioral factors during the life-course may pose considerable and enduring impact on DNAm among aging survivors.

In this study, we hypothesize that DNAm is a mechanism for biological embedding of life experiences including social determinants of health (SDOH) and associated health behaviors. SDOH-related DNAm variations may mediate the association between poor SDOH factors and elevated risk of chronic health conditions (CHC). It has been suggested that epigenetic marks, such as DNAm, could influence the transcription potential of genomic regions and, once changed, could have long-term effects.¹⁶ Furthermore, health behaviors and other environmental factors, such as body mass index (BMI) and smoking which were known to be associated with CHCs, might have a role in this pathway. Systematic epigenome-wide association studies (EWAS) investigating the associations among SDOH factors, DNAm, and CHCs later in life have not yet been conducted in survivors of childhood cancer, even though survivors have much higher burden of morbidity and mortality and tend to be more sensitive to social stressors compared to the general population. To fill this gap, we employed an epigenome-wide approach to systematically identify the DNAm CpGs associated with three key SDOH factors, namely, educational attainment, personal income, and neighborhood SES and physical environment, and evaluated whether SDOH-associated CpGs are also associated with a focused set of pulmonary conditions, related to lung function, which were potentially relevant to smoking. We further applied mediation analysis to explore potential role of DNAm in mediating the effect of each SDOH factor on risk of these conditions.

Results

Characteristics of study population

Study participants included childhood cancer survivors of AA (n=258, median time from diagnosis=25.2 years, interquartile range [IQR]=19.9-32.1 years) and survivors of European ancestry (EA) (n=1,618, median time from diagnosis=27.3 years, IQR=21.1-33.7 years) (**Table 1**). The median age at primary diagnosis was 9.6 (IQR=4.2-14.4) years in AA and 9.0 (IQR=3.8-14.4) years in EA and median age at DNA sampling was 33.9 (IQR=29.4-39.6) years in AA and 35.3 (IQR=30.0-42.1) years in EA. The proportion of female was 53.1% in AA and 40.4% in EA. More than 60% (61.6% in AA and 61.7% in EA) was overweight ($\geq 25\text{kg/m}^2$). Smoking was significantly different with 70.2% of non-smokers in AA and 58.2% of non-smokers in EA ($P<0.0001$). Three SDOH factors were significantly different between AA and EA; specifically, AA had the lower educational attainment ($P<0.0001$), lower personal income ($P<0.0001$), and lived in neighborhoods with poorer socio-economic and physical environment measured by the area deprivation index (ADI) ($P<0.0001$) than EA. Primary cancer diagnoses and cancer treatment information for the study population are also listed in **Table 1**.

Association of DNA methylation sites with SDOH factors

First, EWAS was conducted in EA. The association P -values in $-\log_{10}$ scale were shown for three series of EWAS analyses across the genome in the circos plot (**Figure 1**), with 337 for educational attainment, 39 for personal income, and 24 CpG sites for ADI reached over epigenome-wide significance level with $P<9\times 10^{-8}$, respectively. There were 13 significant DNAm CpGs commonly associated with all three SDOH factors. Those 13 CpGs were mapped to the five genomic regions on chromosomes 2, 3, 5, 9, and 14 harboring genes including *GPR55*, *CLDND1*, *CPOX*, *GPR15*, *AHRR*, *PRRC2B*, and *ELMSAN1*, which are previously known to be

associated with smoking exposures (**Table 2**). After additionally adjusting for BMI and smoking, all 13 CpGs remained statistically significant ($P < 0.05$), but with attenuated effect sizes remaining on average 36.8% for educational attainment (range=30.8%-48.8%), 48.3% for personal income (range=38.5-58.8%), and 43.3% for ADI (range=35.1-61.5%). Given SDOH-CpG associations related to and attenuated by smoking, this finding prompted us to hypothesize that health outcomes related to lung functions may be most likely impacted by the SDOH-associated CpGs.

In the EA model additionally adjusted by BMI and smoking, there were still three epigenome-wide significant associations between educational attainment and cg04180924 (chr3, *CPOX*, $P = 2.0 \times 10^{-8}$), cg04885881 (chr1, intergenic, $P = 1.3 \times 10^{-9}$), cg06359375 (chr22, *HPS4*, $P = 2.0 \times 10^{-8}$). There was one epigenome-wide significant association between personal income and cg04180924 (chr3, *CPOX*, $P = 5.4 \times 10^{-9}$), but no single CpG reached epigenome-wide significant level for ADI. We chose to report the models without adjusting for BMI and smoking as the primary results, however, we applied the condition of $P < 0.05$ in the model additionally adjusted by BMI and smoking to filter against all epigenome-wide significant DNAm CpG sites in the corresponding models without BMI and smoking adjustments. Following this filtering step, we had 130 SDOH-CpG associations including educational attainment (N=88), personal income (N=23), and ADI (N=19). The detailed estimates were provided (**Supplementary Table 1**).

Similarly, EWAS analyses were performed in AA survivors and none of DNAm CpGs reached epigenome-wide significance level ($P < 9 \times 10^{-8}$). However, for comparison purpose, we presented the AA results for 130 SDOH-CpG associations statistically established in EA survivors (**Supplementary Table 2**). Among 130 SDOH-CpG associations, 25 associations were also

validated in AA with $P < 0.05$; specifically, 15 for educational attainments, 8 for personal income, and 3 for ADI.

Association of social determinants of health with pulmonary conditions

The incidence of three common pulmonary conditions were examined among EA and AA survivors using 2 or 3 as the minimum Common Terminology Criteria for Adverse Events (CTCAE)-grade (**Table 1 and Figure 2**). In the unadjusted model with minimum CTCAE-grade set to 2, pulmonary diffusion deficit was significantly higher in AA (25.2% in AA and 18.2% in EA, $P = 0.03$), and restrictive pulmonary deficit was also significantly higher in AA (14.2% in AA and 7.5% in EA, $P = 0.002$), whereas obstructive pulmonary deficit was comparable between the two racial groups (9.8% in AA and 13.1% in EA, $P = 0.21$). In the model adjusting for other covariates, regardless of with or without SDOH factors, race was significantly associated with restrictive pulmonary deficits. However, race was significantly associated with pulmonary diffusion deficits only in the model without adjusting for SDOH factors and became not significant after adjusting for SDOH (**Supplementary Table 3**), suggesting significant contribution of SDOH factors to association between race and pulmonary diffusion deficits.

Methylation mediation effect for social determinants of health on pulmonary conditions

Among all 130 SDOH-CpG associations from EWAS in EA survivors, 61 CpGs (29 educational attainment associated CpG sites, 16 personal income associated CpGs sites, and 16 ADI associated CpG sites) were also significantly associated with risk of pulmonary diffusion deficits after adjusting for multiple comparisons ($P_{FDR} < 0.05$) (**Supplementary Table 4**). In the mediation

analysis, 17 out of 29 educational attainment associated CpGs were identified with significant average causal mediation effects (ACME) after adjusting for multiple comparisons. Using squared pairwise Pearson correlation coefficient r^2 threshold of 0.05, three independent CpGs including cg04180924 (chr3, *CPOX*, mediation=32.9%, $P_{FDR}=0.01$), cg11205006 (chr3, *CPOX*, mediation=32.9%, $P_{FDR}=0.01$), and cg27470486 (chr17, *ACLY*, mediation=8.6%, $P_{FDR}=0.04$) were obtained by top-down pruning the 17 CpGs sorted by estimated ACME in decreasing order. For the final mediation analysis, a combined score (i.e., summation of DNAm levels of the three CpGs with the same direction of association) was used as the mediating variable and 48.9% of mediation effect for educational attainment on pulmonary diffusion deficits was achieved (**Table 3**). Similarly, single significant mediator cg08064403 was found with significant partial mediation effect for personal income (mediation=25.9%, $P<0.001$) and ADI (mediation=24.1%, $P<0.001$) on pulmonary diffusion deficits (**Table 3**).

Methods

Study population

The St. Jude Lifetime Cohort (SJLIFE) study, a retrospective hospital-based cohort study for survivors of childhood cancer with prospective clinical follow-up, has been described elsewhere.^{17,18} Survivors completed demographic and epidemiological questionnaires, received clinical assessments at follow-up visits. A total of 2,052 EA survivors and 370 AA survivors also had DNAm profiling data generated with EPIC array.¹⁵ We subsequently excluded survivors with attained age younger than 25 years old (n=533) to ensure most individuals should have possibly attained the highest level of education (i.e., post-college), and principal components analysis

(PCA) outliers (n=13). The remaining 1,618 EA and 258 AA survivors were included in this current study.

Social determinants of health

SDOH factors for cancer survivors at personal and neighborhood levels were considered and collected at the same time as or before DNA sampling. The personal level factors included educational attainment and personal income measures. Neighborhood social determinants included area SES deprivation as a risk factor. Educational attainment and personal income were collected from SJLIFE surveys. Personal educational attainment was categorized into five levels (below high school, high school or training after high school, some college, college graduate, and post-graduate), and household incomes was categorized into five levels (None, <20, 20K to <40K, 40 to <60K, and ≥60K). We evaluated area deprivation of each survivor using the ADI that is comprised of 17 neighborhood-based SES measures including income, employment, education, and housing status collected in American Community Survey.^{19,20} Census tract and block level of data were created by geocoding the participants' residential addresses using Geographic Identifiers (GEOIDs) from Census Bureau, then geocoded data was linked to national ADI file. Technically, these measures present SES and physical environment by census blocks in the USA, and each census block receives a percentile ranking with minimum disadvantage in the 1st percentile and maximum disadvantage in the 100th percentile. For ADI, we considered >75th percentile, 40 to 75 percentile, and <40 percentile as high, moderate, and low area SES deprivation, respectively.

Chronic health conditions

We will focus on the following three common CHC related to lung function: obstructive pulmonary deficit, pulmonary difficult deficit, and restrictive pulmonary deficit. Clinical evaluations of these common conditions have been documented in previous publications for SJLIFE cohort.^{17,18} CHCs with severity grading based on CTCAE version 4.0: 0 (no problem), 1 (mild), 2 (moderate), 3 (severe/disabling), 4 (life-threatening), or 5 (death).¹⁸ Only incident CHC with grade ≥ 2 and occurred after blood draw for DNAm profiling were considered as an adverse event in the current study.

DNA methylation profiling

Blood samples collected from each participant at their follow-up visits were banked in the St. Jude Biorepository. Methylation profiling were performed using the Infinium[®] MethylationEPIC BeadChip. Genomic DNA (250ng per sample) were treated with bisulfite using the Zymo EZ DNA Methylation Kit under the following thermos-cycling conditions: 16 cycles: 95°C for 30 sec, 50°C for 1 hour. Following bisulfite treatment, DNA samples were desulphonated, column purified, then eluted using 12 ul of elution buffer (Zymo Research Corp.). Treating DNA with sodium bisulfite selectively converted cytosine to uracil but preserves 5-methylcytosine, which was immune from deamination. Bisulfite-converted DNA (4 μ l) was then processed by following the Illumina Infinium Methylation Assay protocol, which included hybridization to MethylationEPIC BeadChips, single base extension assay, and staining and scanning using the Illumina HiScan system.

DNA methylation bioinformatic analysis

Methylation raw intensity data were analyzed in R (version 3.6.1) using minfi package.²¹ The processed methylation data was described as a β -value, which is a continuous variable ranging between 0 (no methylation) and 1 (full methylation). In any sample, a probe with a detection p value more than 0.01 was assigned missing status. Any sample or probe with more than 5% missing values were excluded from downstream analysis. Non-specific or cross-reactive probes or probes with SNPs nearby the CpG site were also be excluded. The methylation data were normalized by quantile normalization method implemented in minfi package.²¹ M-value (logit transformation of β -value) were calculated and subsequently used in regression analyses.²² The epigenome-wide methylation scores were used to impute six lymphocyte cell subtype proportions (lymphocyte, monocyte, and granulocyte, CD4, natural killer, and B cells).^{23,24} A PCA were performed to quantify latent structures or batch effects in the data. We use the array annotations provided by Illumina to map probes to their corresponding genes.

Statistical analysis

The EWAS between methylation level at each CpG-site (dependent variable, continuous) and each SDOH factor (independent variable, categorical) were evaluated using multiple linear regression with covariate adjustments including sex, attained age, race, cancer treatment exposures, lymphocyte subtype proportions, significant genetic principal components and methylation principal components (model 1). Health behaviors (BMI and smoking) were additionally adjusted to evaluate the contribution of each SDOH factor on methylation level

independent of health behaviors (model 2). However, because health behaviors are intermediate variables and potentially on the causal pathway from SDOH factors to DNAm, we would avoid the over adjustment²⁵ and treated epigenome-wide significant results from model 1 as the primary results with a post filtering step with the criteria of $P < 0.05$ in model 2 to derive more credible SDOH-associated CpGs considering the effect did not completely disappear in model 2. We used R package CpGassoc²⁶ for the linear regression analysis and used $P < 9 \times 10^{-8}$ as the epigenome-wide significance threshold.²⁷ EWAS results were visualized with circos plot (Circos v.0.69).²⁸

Multivariable logistic regression models were used to assess the associations between the adjusted M-value for each SDOH-associated CpG (by the same set of covariates specified in aforementioned EWAS model) and each of pulmonary condition (binary outcome) with covariates including sex, age, BMI, smoking, and radiation to the chest (a treatment exposure significantly associated with pulmonary conditions). Only CpG-CHC associations with the significant false discovery rate (FDR) adjusted p -value ($P_{\text{FDR}} < 0.05$) were regarded as statistical significance and were advanced as candidate CpGs for the downstream mediation analysis.

In the mediation analysis, each of the candidate CpGs was considered as a hypothesized mediating variable, each SDOH factor as the exposure, status of each pulmonary condition (binary) as the outcome, and adjusted for sex, age, BMI, smoking, and radiation to the chest. Individually significant mediator CpGs will be checked for pair-wise correlations. All independent significant mediator CpGs (squared Pearson correlation coefficient $r^2 < 0.05$) were combined into a methylation risk score for a final mediation analysis. Mediation analysis was

carried out using Mediation R package.²⁹ All statistical analyses were performed by using R.3.6.3³⁰ or SAS 9.4 (SAS Institute Inc., Cary, NC, USA) and all statistical tests were two-sided.

Discussion

To the best of our knowledge, this is the first social epigenomic study conducted among survivors of childhood cancer. We found 130 SDOH-CpG associations for educational attainment, personal income, or ADI at epigenome-wide significance level. However, there were an order of magnitude fewer number of CpGs than those associated with cancer treatment exposures,¹⁵ suggesting that the SDOH factors may have smaller effect sizes compared to intensive cancer treatment modalities. Among all three SDOH factors that were examined, educational attainment seems to be the most informative factor with the largest number of significant CpG associations. Most of the CpGs were mapped to genes that were known to have DNAm associated with smoking exposure, suggesting that social and behavioral exposures to lower educational attainment, lower income and higher SES disadvantageous neighborhood resemble the effect of tobacco use. More interestingly, there were also some robust findings not related to smoking exposures. For example, cg06359375 was uniquely associated with educational attainment (but neither personal income nor area deprivation index) at epigenome-wide significant level with no attenuation for the effect size after adjusting for BMI and smoking. This CpG is mapped to *HPS4* gene, which was previously reported to be associated with cognitive function.³¹

Although biological pathways from which socio-economic vulnerability influences health outcomes in survivors of childhood cancer have not been determined, according to a biological

embedding of childhood adversity model, childhood stress can get programmed into macrophage through epigenetic markings, post-translational modifications, and tissue remodeling. This could cause the immune cells to mount excessive inflammatory to microbial challenges and be insensitivity to inhibitory hormonal signals, resulting in a chronic inflammatory state in the body.³² Moreover, several studies have supported that SES was biologically embedded by showing that low SES across the life course has been associated with a blunted pattern of diurnal cortisol production,³³ higher level of allostatic load,³³ increased inflammatory activity,³⁴⁻³⁶ and higher pathogen burden.^{37,38}

EWAS findings suggest potential implications of SDOH-associated CpGs in risk of three common pulmonary diseases that were clinically-assessed during SJLIFE follow-up visits. We found AA survivors had significantly higher incidence of pulmonary diffusion deficit or restrictive pulmonary deficit than EA survivors. In contrast, there is no racial disparity for incidence of obstructive pulmonary deficit. Considering AA survivors had worse socio-economic exposures than EA based on all three key factors (i.e., lower educational attainment, lower personal income, and higher national rank of ADI), we continued to examine whether SDOH can account for, to some extent, the observed higher incidence of pulmonary diseases. We observed two distinct patterns; the racial disparity in pulmonary diffusion deficits disappeared after adjusting for SDOH factors suggesting significant contribution of SDOH factors to racial disparity in pulmonary diffusion deficits, whereas the racial disparity in restrictive pulmonary deficits remained significant with little change.

To illuminate potential biological mechanism for the SDOH effect on pulmonary diffusion deficit, we evaluated the association between SDOH-associated CpGs and incidence

of pulmonary diffusion deficit and further performed mediation analysis for a subset of CpGs (mediator) that are associated with both SDOH (exposure) and pulmonary diffusion deficit (outcome). There were three CpG collectively contributed nearly 50% mediation effect of educational attainment, and a single mediator CpG had partial mediation of personal income and ADI on pulmonary diffusion deficit. In contrast, we did not find any of these SDOH associated CpGs were associated with either obstructive pulmonary deficit or restrictive pulmonary deficit.

Our study has the following limitations; First, the sample size of the AA population was small for identification of race specific epigenetic association with SDOH factors, but 20% (26/130) epigenome-wide significant CpGs among EA survivors were validated using AA survivors ($P < 0.05$). Second, the analysis was based on cross-sectional study, so there was no clearly defined temporal association to establish the causality. Ideally, DNAm measured at two time points for the same set of survivors with several years apart would allow us to more rigorously analyze the effect of baseline SDOH factors (or change of SDOH) on the changes of DNAm between the two consecutive time points and also evaluate the temporal association between DNAm measured at the earlier time point and incidence CHC occurred at later time point. Third, we attempted to take advantage of the existing whole-genome sequencing data to search for methylation quantitative trait loci (meQTL) but we did not find any strong meQTL for the top genomic region on chromosome 3 that can be used in mendelian randomization³⁹ for causal inference.

In summary, the first social epigenomic study in childhood cancer survivors identified striking DNAm signatures associated with multiple SDOH factors, with many epigenome-wide

significant CpG sites that mimic the effect of smoking exposure. By taking a social epigenomic approach, we also demonstrated an exemplified racial health disparity in pulmonary diffusion deficit between AA and EA survivors, and illuminated DNAm as potential mediators for SDOH factors which contributed to the disparity.

References

1. Bhatia, S. *et al.* Childhood cancer survivorship research in minority populations: A position paper from the Childhood Cancer Survivor Study. *Cancer* **122**, 2426-39 (2016).
2. Castellino, S.M. *et al.* Minority adult survivors of childhood cancer: a comparison of long-term outcomes, health care utilization, and health-related behaviors from the childhood cancer survivor study. *J Clin Oncol* **23**, 6499-507 (2005).
3. Dixon, S.B. *et al.* Racial and ethnic disparities in neurocognitive, emotional, and quality-of-life outcomes in survivors of childhood cancer: A report from the Childhood Cancer Survivor Study. *Cancer* **125**, 3666-3677 (2019).
4. Liu, Q. *et al.* Racial/Ethnic Differences in Adverse Outcomes Among Childhood Cancer Survivors: The Childhood Cancer Survivor Study. *J Clin Oncol* **34**, 1634-43 (2016).
5. Cunliffe, V.T. The epigenetic impacts of social stress: how does social adversity become biologically embedded? *Epigenomics* **8**, 1653-1669 (2016).
6. Mancilla, V.J. *et al.* Understanding the Interplay Between Health Disparities and Epigenomics. *Front Genet* **11**, 903 (2020).
7. Borghol, N. *et al.* Associations with early-life socio-economic position in adult DNA methylation. *Int J Epidemiol* **41**, 62-74 (2012).
8. Lam, L.L. *et al.* Factors underlying variable DNA methylation in a human community cohort. *Proc Natl Acad Sci U S A* **109 Suppl 2**, 17253-60 (2012).
9. McGuinness, D. *et al.* Socio-economic status is associated with epigenetic differences in the pSoBid cohort. *Int J Epidemiol* **41**, 151-60 (2012).

10. Stringhini, S. *et al.* Life-course socioeconomic status and DNA methylation of genes regulating inflammation. *Int J Epidemiol* **44**, 1320-30 (2015).
11. Needham, B.L. *et al.* Life course socioeconomic status and DNA methylation in genes related to stress reactivity and inflammation: The multi-ethnic study of atherosclerosis. *Epigenetics* **10**, 958-69 (2015).
12. Karlsson Linnér, R. *et al.* An epigenome-wide association study meta-analysis of educational attainment. *Mol Psychiatry* **22**, 1680-1690 (2017).
13. van Dongen, J. *et al.* DNA methylation signatures of educational attainment. *NPJ Sci Learn* **3**, 7 (2018).
14. Joehanes, R. *et al.* Epigenetic Signatures of Cigarette Smoking. *Circ Cardiovasc Genet* **9**, 436-447 (2016).
15. Song, N. *et al.* Persistent Variations of Blood DNA Methylation Associated with Treatment Exposures and Risk for Cardiometabolic Outcomes among Long-term Survivors of Childhood Cancer: A Report from the St. Jude Lifetime Cohort. *medRxiv*, 2020.09.10.20192393 (2020).
16. Jaenisch, R. & Bird, A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat Genet* **33 Suppl**, 245-54 (2003).
17. Hudson, M.M. *et al.* Prospective medical assessment of adults surviving childhood cancer: study design, cohort characteristics, and feasibility of the St. Jude Lifetime Cohort study. *Pediatr Blood Cancer* **56**, 825-36 (2011).

18. Hudson, M.M. *et al.* Approach for Classification and Severity Grading of Long-term and Late-Onset Health Events among Childhood Cancer Survivors in the St. Jude Lifetime Cohort. *Cancer Epidemiol Biomarkers Prev* **26**, 666-674 (2017).
19. Kind, A.J. *et al.* Neighborhood socioeconomic disadvantage and 30-day rehospitalization: a retrospective cohort study. *Ann Intern Med* **161**, 765-74 (2014).
20. Kind, A.J.H. & Buckingham, W.R. Making Neighborhood-Disadvantage Metrics Accessible - The Neighborhood Atlas. *N Engl J Med* **378**, 2456-2458 (2018).
21. Aryee, M.J. *et al.* Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* **30**, 1363-9 (2014).
22. Du, P. *et al.* Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics* **11**, 587 (2010).
23. Houseman, E.A. *et al.* DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics* **13**, 86 (2012).
24. Jaffe, A.E. & Irizarry, R.A. Accounting for cellular heterogeneity is critical in epigenome-wide association studies. *Genome Biol* **15**, R31 (2014).
25. Schisterman, E.F., Cole, S.R. & Platt, R.W. Overadjustment bias and unnecessary adjustment in epidemiologic studies. *Epidemiology* **20**, 488-95 (2009).
26. Barfield, R.T., Kilaru, V., Smith, A.K. & Conneely, K.N. CpGassoc: an R function for analysis of DNA methylation microarray data. *Bioinformatics* **28**, 1280-1 (2012).
27. Mansell, G. *et al.* Guidance for DNA methylation studies: statistical insights from the Illumina EPIC array. *BMC Genomics* **20**, 366 (2019).

28. Krzywinski, M. *et al.* Circos: an information aesthetic for comparative genomics. *Genome Res* **19**, 1639-45 (2009).
29. Tingley, D., Yamamoto, T., Hirose, K., Keele, L. & Imai, K. mediation: R Package for Causal Mediation Analysis. *2014* **59**, 38 (2014).
30. Team., R.D.C. R: A language and environment for statistical computing. (R Foundation for Statistical Computing., Vienna, Austria, 2010).
31. Kuratomi, G. *et al.* Association of the Hermansky-Pudlak syndrome type 4 (HPS4) gene variants with cognitive function in patients with schizophrenia and healthy subjects. *BMC Psychiatry* **13**, 276 (2013).
32. Miller, G.E., Chen, E. & Parker, K.J. Psychological stress in childhood and susceptibility to the chronic diseases of aging: moving toward a model of behavioral and biological mechanisms. *Psychol Bull* **137**, 959-97 (2011).
33. Dowd, J.B., Simanek, A.M. & Aiello, A.E. Socio-economic status, cortisol and allostatic load: a review of the literature. *Int J Epidemiol* **38**, 1297-309 (2009).
34. Ranjit, N. *et al.* Socioeconomic position, race/ethnicity, and inflammation in the multi-ethnic study of atherosclerosis. *Circulation* **116**, 2383-90 (2007).
35. Loucks, E.B. *et al.* Life course socioeconomic position is associated with inflammatory markers: the Framingham Offspring Study. *Soc Sci Med* **71**, 187-95 (2010).
36. Carroll, J.E., Cohen, S. & Marsland, A.L. Early childhood socioeconomic status is associated with circulating interleukin-6 among mid-life adults. *Brain Behav Immun* **25**, 1468-74 (2011).

37. Steptoe, A. *et al.* Socioeconomic status, pathogen burden and cardiovascular disease risk. *Heart* **93**, 1567-70 (2007).
38. Dowd, J.B., Zajacova, A. & Aiello, A. Early origins of health disparities: burden of infection, health, and socioeconomic status in U.S. children. *Soc Sci Med* **68**, 699-707 (2009).
39. Teumer, A. Common Methods for Performing Mendelian Randomization. *Front Cardiovasc Med* **5**, 51 (2018).

Table 1. Characteristics of study participants

	African-Ancestry (N=258)		European-Ancestry (N=1,618)		P ^a
	N	(%)	N	(%)	
General characteristics					
Sex					0.05
Female	137	(53.1%)	654	(40.4%)	
Male	121	(46.9%)	761	(47.0%)	
Body mass index (BMI), kg/m ²					0.66
<18.5	9	(3.5%)	47	(2.9%)	
18.5-<25.0	67	(26.0%)	469	(29.0%)	
≥25.0	159	(61.6%)	999	(61.7%)	
Unknown	23	(8.9%)	103	(6.4%)	
Median (IQR)	27.8	(23.3-33.0)	27.2	(23.5-32.3)	0.27
Smoking					<0.0001
Non-smokers	181	(70.2%)	941	(58.2%)	
Past smokers	19	(7.4%)	278	(17.2%)	
Current smokers	45	(17.4%)	355	(21.9%)	
Unknown	13	(5.0%)	44	(2.7%)	
Age at primary diagnosis, years					
Median (IQR)	9.6	(4.2-14.4)	9.0	(3.8-14.4)	0.85
Age at DNA sampling, years					
Median (IQR)	33.9	(29.4-39.6)	35.3	(30.3-42.1)	0.03
Time from diagnosis to DNA sampling, years					
Median (IQR)	25.2	(19.9-32.1)	27.3	(21.1-33.7)	0.03
Social determinants of health (SDOH)					
Educational attainment					<0.0001
Below high school	39	(15.1%)	137	(8.5%)	
High school or training after high school	50	(19.4%)	277	(17.1%)	
College	84	(32.6%)	401	(24.8%)	
College graduate	49	(19.0%)	478	(29.5%)	
Post-graduate	12	(4.7%)	179	(11.1%)	
Unknown	24	(9.3%)	146	(9.0%)	
Personal income, dollars					<0.0001
None	38	(14.7%)	143	(8.8%)	
<20K	113	(43.8%)	502	(31.0%)	
20K-<40K	53	(20.5%)	442	(27.3%)	
40K-<60K	19	(7.4%)	256	(15.8%)	
≥60K	35	(13.6%)	275	(17.0%)	
Area deprivation index (ADI), national rank					
Median (IQR)	80.0	(56.0-91.5)	56.0	(36.0-76.0)	<0.0001
Clinical characteristics					
Diagnosis					<0.0001
Leukemia	50	(19.4%)	548	(33.9%)	
Lymphoma	57	(22.1%)	639	(39.5%)	
CNS tumors	27	(10.5%)	136	(8.4%)	
Others	124	(48.1%)	525	(32.4%)	
Chemotherapy					
Alkylating agents	146	(56.6%)	973	(60.1%)	0.28
Anthracyclines	122	(47.3%)	931	(57.5%)	0.002
Corticosteroids	85	(32.9%)	791	(48.9%)	<0.0001
Epipodophyllotoxins	63	(24.4%)	501	(31.0%)	0.03

Platinum	38	(14.7%)	151	(9.3%)	0.008
Vincristine	158	(61.2%)	1,159	(71.6%)	0.0007
Radiation therapy (RT)					
Brain RT	72	(27.9%)	525	(32.4%)	0.15
Chest RT	81	(31.4%)	513	(31.7%)	0.92
Abdomen/Pelvis RT	76	(29.5%)	391	(24.2%)	0.07
Common pulmonary conditions					
Obstructive pulmonary deficit					
Grade ≥ 2 (N=1,397)	18	(9.8)	159	(13.1)	0.21
Grade ≥ 3 (N=1,446)	10	(5.1)	50	(4.0)	0.47
Pulmonary diffusion deficits					
Grade ≥ 2 (N=1,315)	42	(25.2)	209	(18.2)	0.03
Grade ≥ 3 (N=1,445)	5	(2.6)	22	(1.8)	0.43
Restrictive pulmonary deficit					
Grade ≥ 2 (N=1,404)	26	(14.2)	91	(7.5)	0.002
Grade ≥ 3 (N=1,446)	12	(6.2)	29	(2.3)	0.002

Abbreviations: BMI (body mass index), IQR (interquartile range), area deprivation index (ADI), CNS (central nervous system), RT (radiation therapy)

^aChi-square test for categorical variables, T-test for continuous variables, and Wilcoxon rank sum test for ranked variables.

Table 2. DNA methylation CpG sites that are significantly associated with all three social determinants of health factors

CpG label	Chr	Position	HGNC gene	Educational attainment				Personal income				Area deprivation index			
				Effect size	SE	P	% Effect remaining ^a	Effect size	SE	P	% Effect remaining ^a	Effect size	SE	P	% Effect remaining ^a
cg01731783	chr14	73745084	<i>ELMSAN1</i>	0.0181	0.0024	1.9E-13	30.8%	0.0133	0.0022	2.2E-09	52.3%	-0.0007	0.0001	3.1E-08	44.9%
cg08840017	chr2	230944894	<i>GPR55</i>	0.0345	0.0049	2.9E-12	43.9%	0.0258	0.0044	7.2E-09	55.9%	-0.0013	0.0002	4.0E-08	58.7%
cg00385142	chr3	98517073	<i>CLDND1</i>	0.0389	0.0041	1.8E-20	48.8%	0.0246	0.0038	1.0E-10	58.8%	-0.0015	0.0002	7.1E-14	61.5%
cg18754985	chr3	98518905	<i>CLDND1</i>	0.0536	0.0054	7.9E-23	32.4%	0.0373	0.0049	4.6E-14	50.8%	-0.0017	0.0003	5.2E-10	39.1%
cg08064403	chr3	98521413	<i>CLDND1</i> ; <i>CPOX</i> ;	0.1034	0.0092	3.4E-28	32.2%	0.0695	0.0084	2.6E-16	49.3%	-0.0033	0.0005	1.3E-12	41.4%
			<i>RP11-227H4.5</i> <i>CLDND1</i> ;												
cg05659611	chr3	98521779	<i>CPOX</i> ; <i>RP11-227H4.5</i>	0.0647	0.0055	5.0E-31	40.6%	0.0375	0.0051	1.9E-13	50.1%	-0.0024	0.0003	1.3E-17	51.6%
cg19859270	chr3	98532449	<i>CPOX</i> ; <i>GPR15</i>	0.1086	0.0083	4.5E-37	34.1%	0.0671	0.0077	8.7E-18	43.3%	-0.0032	0.0004	4.4E-14	35.1%
cg04180924	chr3	98553219	<i>CPOX</i>	0.0997	0.0078	1.9E-35	39.0%	0.0689	0.0072	3.8E-21	53.3%	-0.0028	0.0004	4.4E-12	38.3%
cg00010201	chr3	98568261	<i>CPOX</i>	0.0806	0.0076	1.4E-25	36.1%	0.0488	0.0069	2.1E-12	44.4%	-0.0023	0.0004	1.3E-09	31.4%
cg02978227	chr3	98573182	<i>CPOX</i>	0.0798	0.0066	3.5E-32	30.8%	0.0481	0.0061	8.1E-15	40.7%	-0.0023	0.0003	1.8E-12	35.9%
cg02657160	chr3	98592218	<i>CPOX</i>	0.0547	0.0048	8.9E-29	36.4%	0.0366	0.0044	2.0E-16	48.3%	-0.0019	0.0002	7.9E-15	48.9%
cg05575921	chr5	373262	<i>AHRR</i>	0.1361	0.0104	8.1E-37	29.4%	0.0868	0.0097	1.3E-18	42.0%	-0.0038	0.0005	2.5E-13	35.9%
cg26768182	chr9	131397291	<i>PRRC2B</i>	0.0613	0.0057	4.6E-26	44.0%	0.0286	0.0053	8.0E-08	38.5%	-0.0015	0.0003	5.1E-08	40.0%
Combined							36.8%				48.3%				43.3%

Abbreviations: Chr (chromosome), HGNC (The HUGO genome nomenclature committee), SE (standard error), BMI (body mass index)

^a% Effect remaining was calculated by effect size of model 2 additionally adjusted for smoking and BMI divided by original effect size of model 1.

Table 3. DNA methylation partially mediates the association between social determinants of health factors and risk of pulmonary diffusion deficits

Social determinants of Health	CpG	Chromosome	Position	HGNC gene	ACME	ADE	Total effect ^a	% Mediation ^b	P _{FDR}
Educational attainment	cg04180924	chr3	98,553,219	<i>CPOX</i>	-0.0137	-0.0280	-0.0417	32.9%	0.01
	cg11205006	chr22	26,479,532	<i>HPS4</i>	-0.0058	-0.0246	-0.0304	19.0%	0.02
	cg27470486	chr17	41,917,434	<i>ACLY</i>	-0.0029	-0.0307	-0.0336	8.6%	0.04
	Combined Score ^a				-0.0214	-0.0224	-0.0438	48.9%	0.002
Personal income	cg08064403	chr3	98,521,413	<i>CLDND1;CPOX;RP11-227H4.5</i>	-0.0101	-0.0290	-0.0392	25.9%	<0.001
Area deprivation index	cg08064403	chr3	98,521,413	<i>CLDND1;CPOX;RP11-227H4.5</i>	0.0002	0.0007	0.0009	24.1%	<0.001

Abbreviations: Chr (chromosome), HGNC (The HUGO genome nomenclature committee), ACME (average causal mediation effect), ADE (average direct effect), FDR (false-discovery rate)

^aCalculated by summing ACME and ADE.

^bCalculated by ACME divided by total effect (the sum of ACME and ADE).

Figure 1. Circos plot for epigenome-wide association studies (EWAS) of three social determinants of health factors. Outer circle (red): EWAS for educational attainment; middle circle (green): EWAS for personal income; inner circle (blue): EWAS for area deprivation index. Each dot depicts $-\log_{10}$ p value for each DNAm CpG site mapped to a chromosome location along the genome. The black lines indicate the epigenome-wide significance level ($P=9 \times 10^{-8}$).

Figure 2. Incidence of common pulmonary conditions among survivors of African ancestry and European ancestry