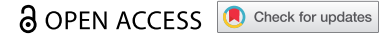




RESEARCH PAPER



Blood DNA methylation signatures are associated with social determinants of health among survivors of childhood cancer

Nan Song ^{a,b}, Jin-Ah Sim^{a,c,*}, Qian Dong^{a,*}, Yinan Zheng^d, Lifang Hou^d, Zhenghong Li^a, Chia-Wei Hsu^e, Haitao Pan^{a,e}, Heather Mulder^{a,f}, John Easton^{a,f}, Emily Walker^{a,g}, Geoffrey Neale^{a,g}, Carmen L. Wilson^a, Kirsten K. Ness^a, Kevin R. Krull^{a,h}, Deo Kumar Srivastava^{a,e}, Yutaka Yasui^a, Jinghui Zhang^{a,f}, Melissa M. Hudson^{a,i}, Leslie L. Robison^a, I-Chan Huang^{a,*}, and Zhaoming Wang ^{a,f,#}

^aDepartment of Epidemiology and Cancer Control, St. Jude Children's Research Hospital, Memphis, TN, USA; ^bDepartment of Pharmacy, College of Pharmacy, Chungbuk National University, Cheongju, Chungcheongbuk-do, Korea; ^cSchool of Ai Convergence, Hallym University, Chuncheon, Gangwon-do, Korea; ^dDepartment of Preventive Medicine, Northwestern University, Chicago, IL, USA; ^eDepartment of Biostatistics, St. Jude Children's Research Hospital, Memphis, TN, USA; ^fDepartment of Computational Biology, St. Jude Children's Research Hospital, Memphis, TN, USA; ^gSt. Jude Children's Research Hospital, Hartwell Center, Memphis, TN, USA; ^hDepartment of Psychology, St. Jude Children's Research Hospital, Memphis, TN, USA; ⁱDepartment of Oncology, St. Jude Children's Research Hospital, Memphis, TN, USA

ABSTRACT

Social epigenomics is an emerging field in which social scientist collaborate with computational biologists, especially epigeneticists, to address the underlying pathway for biological embedding of life experiences. This social epigenomics study included long-term childhood cancer survivors enrolled in the St. Jude Lifetime Cohort. DNA methylation (DNAm) data were generated using the Illumina EPIC BeadChip, and three social determinants of health (SDOH) factors were assessed: self-reported educational attainment, personal income, and an area deprivation index based on census tract data. An epigenome-wide association study (EWAS) was performed to evaluate the relation between DNAm at each 5'-cytosine-phosphate-guanine-3' (CpG) site and each SDOH factor based on multivariable linear regression models stratified by ancestry (European ancestry, $n = 1,618$; African ancestry, $n = 258$). EWAS among survivors of European ancestry identified 130 epigenome-wide significant SDOH–CpG associations ($P < 9 \times 10^{-8}$), 25 of which were validated in survivors of African ancestry ($P < 0.05$). Thirteen CpGs were associated with all three SDOH factors and resided at pleiotropic loci in cigarette smoking-related genes (e.g., *CLDND1* and *CPOX*). After accounting for smoking and body mass index, these associations remained significant with attenuated effect sizes. Seven of 13 CpGs were associated with gene expression level based on 57 subsamples with blood RNA sequencing data available. In conclusion, DNAm signatures, many resembling the effect of tobacco use, were associated with SDOH factors among survivors of childhood cancer, thereby suggesting that biologically distal SDOH factors influence health behaviours or related factors, the epigenome, and subsequently survivors' health.

ARTICLE HISTORY

Received 24 May 2021
Revised 10 January 2022
Accepted 13 January 2022



KEYWORDS

DNA methylation; social determinants of health; childhood cancer; survivorship; epigenome-wide association study

Introduction


Social and behavioural epigenomics is an emerging transdisciplinary field that investigates how social determinants of health (SDOH) and health behaviours modulate the human epigenome and thus influence health and wellness [1,2]. Among various epigenetic modifications, DNA methylation (DNAm) has been the most widely studied; multiple groups have reported significant associations between DNAm and socio-economic status (SES)

[1,3–7], educational attainment [8,9], and health behaviours [10]. These findings support the hypothesis that DNAm is a potential underlying mechanism for biological embedding of life experiences, including physiologically distal SDOH factors and more directly relevant health behaviours. Previous studies have identified variations of DNAm across the methylome related to SDOH factors that resemble variations seen among persons who engage in smoking, risky or

CONTACT Zhaoming Wang  Zhaoming.Wang@stjude.org  Department of Epidemiology and Cancer Control, St. Jude Children's Research Hospital, 262 Danny Thomas Place, MS 735, Memphis, TN 38105

*These authors contributed equally as co-first authors.

#These authors contributed equally as co-senior authors.

 Supplemental data for this article can be accessed [here](#).

© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

heavy alcohol consumption, or who are obese [5,8,9,11,12]. Moreover, findings demonstrate that associations between some of these DNAm variations and educational attainment remain significant after accounting for known risk factors such as smoking, thereby suggesting traces of other exposures for less-educated people [8,9]

Epigenetic signatures provide insight into the mechanistic underpinnings of the link between SES and chronic disease risk [6]. Heritability analyses have shown an approximately 18% variance of DNAm between monozygotic and dizygotic twins, but this difference decreased as the twins aged, implying that non-genetic factors, including SES, might influence DNAm [13,14]. A recent study showed persistent epigenetic ageing and differentially methylated loci among adolescent/young adult Hodgkin lymphoma survivors compared to their unaffected twins [15].

A recent epigenetic analysis of peripheral blood samples from long-term survivors of childhood cancer participating in the St. Jude Lifetime Cohort Study (SJLIFE) found that specific cancer treatments are associated with variations in DNAm decades after exposure [16,17]. Moreover, a subset of treatment-associated DNAm 5'-cytosine-phosphate-guanine-3' (CpG) sites was also significantly associated with cardiometabolic risk, and hence potentially mediated treatment-related toxicities leading to chronic health conditions [16]. Although intensive treatment exposures would be expected to leave persistent epigenetic signatures within a survivor's methylome, socio-economic and behavioural factors during the person's life-course may also have an enduring impact on his/her DNAm.

To our knowledge, no social epigenomic studies investigated the links between SDOH factors and potential epigenetic modifications in survivors of childhood cancer. To fill this gap, we employed a comprehensive epigenome-wide approach to identify DNAm CpGs associated with three SDOH factors (i.e., educational attainment, personal income, and area SES deprivation). These factors may influence the health disparity of childhood cancer survivors who have different racial backgrounds [18] and were taken into consideration based on the data availability in the SJLIFE survey and missingness of available variables. In addition, we annotated these CpGs with

associated traits (exposures or outcomes) catalogued in the EWAS Atlas [19].

Methods

Study population

SJLIFE is a retrospective hospital-based cohort study of 5-year survivors of childhood cancer with prospective clinical follow-up [20]. During periodic follow-up visits, survivors reported social/demographic information based on standard questionnaires and underwent comprehensive, systematic clinical assessments. Among the SJLIFE participants, epigenome-wide DNAm profiling data were available for 2,052 survivors of European ancestry and 370 survivors of African ancestry. For the current report, we excluded survivors younger than 25 years of age ($n = 533$) who may have not yet completed their education (e.g., college or post-graduate degree). In addition, we excluded outliers from principal components analysis (PCA) of genotypes ($n = 13$) to ensure a relatively homogenous population substructure in the analytic data set. The remaining 1,876 survivors (1,618 of European ancestry and 258 of African ancestry) were included in the final analysis. All survivors were deemed free of childhood cancer (survival ≥ 10 years from cancer diagnosis) and fewer than 0.5% of SJLIFE survivors experienced subsequent haematologic malignancies, as of the most last follow-up [21].

Social determinants of health

The SDOH factors of cancer survivors at personal and neighbourhood levels were collected before or at the same time as the blood draw for DNAm analysis. The personal-level SDOH factors included educational attainment and personal annual income as reported in 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 personal annual incomes was categorized into five levels (none, $< \$20,000$, $\$20,000$ to $\$40,000$, $\$40,000$ to $\$60,000$, and $\geq \$60,000$ [U.S. dollars]). Neighbourhood social determinants included area SES deprivation

for each survivor by using the Area Deprivation Index (ADI), which consists of 17 neighbourhood-based SES measures, including income, employment, education, and housing status, mainly collected from the American Community Survey [22,23]. Participants' residential addresses were geocoded using geographic identifiers from the U.S. Census Bureau, which were then linked to the national ADI file. These measures present disadvantaged SES and physical environment by census blocks in the U.S.A., and each census block receives a percentile ranking, with minimum disadvantage in the first percentile and maximum disadvantage in the 100th percentile. For ADI, we considered >75th percentile, 40th to 75th percentile, and <40th percentile as high, moderate, and low SES deprivation area, respectively.

BMI and smoking

At each follow-up examination, the height and weight of the SJLIFE participant was measured. BMI of survivors was calculated as weight (kg) divided by height (m) squared (kg/m^2). According to the standardized classification of the World Health Organization, BMI was categorized as <18.5, 18.5 to <25.0, 25.0 to <30.0 and $\geq 30.0 \text{ kg}/\text{m}^2$. BMI was considered a health behaviour-related factor [24,25]. Smoking status was assessed based on self-reported questionnaires at each follow-up visit and categorized as non-smoker versus former or current smoker. BMI and smoking status at the same or closest date of blood draw for DNAm analysis were used in the current analysis.

DNA methylation profiling and analysis

DNAm data were generated using MethylationEPIC BeadChip and analysed as previously described [16,26]. In brief, methylation raw intensity data were analysed in R (version 3.6.1) using the minfi package [27]. M-values (logit transformation of β -value) were calculated and subsequently used as the dependent variable of regression analyses [28]. Houseman's method was used to estimate leukocyte cell subtype proportions (monocyte, granulocyte, CD4 + T cells, CD8 + T cells, natural killer cells, and B cells) [29,30]. A PCA of DNAm data was

performed to quantify latent structures or batch effects (**Supplementary Figure 1**). We used the array annotations (GRCh38 version) provided by Illumina (San Diego, CA) to map probes to their corresponding genes.

RNA-sequence profiling and data processing

All RNA-sequencing libraries underwent 151-cycle paired-end sequencing on the Illumina NovaSeq 6000 System. Sequencing reads were mapped to the GRCh38 build using STAR [31]. Read counts by gene were generated using HTSeq [32], and genes with read counts fewer than 10 were excluded. Transcript per million (TPM) reads were further calculated, and $\log_2(\text{TPM}+0.01)$ values were quantile-normalized using Limma [33]. The processed gene-by-sample data matrices were used to quantify the gene expression level in further statistical analyses.

Statistical analyses

To illustrate the rationale underlying our modelling approaches, we have provided a conceptual framework for the social epigenomics study in survivors of childhood cancer (**Supplementary Figure 2**). We hypothesized poor SDOH as 'upstream exposures' and risky health behaviours or related factors as 'downstream exposures' that influence epigenetic mechanisms (e.g., DNAm) and regulate gene expression, which ultimately affects health outcomes. Survivors of European ancestry ($n = 1,618$) and those of African ancestry ($n = 258$) were analysed separately, with the former designated as the primary dataset and the latter as the exploratory dataset due to its limited sample size. The EWAS analysis for the association between methylation level at each CpG site (dependent variable) and each SDOH factor (independent variable) was conducted using multiple linear regressions with covariate adjustments including sex, attained age, specific cancer treatment exposures (chemotherapy agents and radiation therapy sites), leukocyte-subtype proportions, significant genetic principal components based on genotypes derived from existing whole-genome sequencing [34] and methylation principal components (model 1). Health behaviours and related

factors (smoking and BMI) were also adjusted to evaluate the contribution of each SDOH factor on the DNAm level independent of health behaviours and related factors (model 2). However, because health behaviours and related factors are potentially intermediate variables and on the causal path from SDOH factors to DNAm, to avoid over adjustment [35], we started with epigenome-wide significant results from model 1 as the primary results but retained only those CpGs satisfying $P < 0.05$ after Bonferroni correction in model 2. We further conducted a meta-analysis of SDOH EWAS among survivors of African ancestry and those of European ancestry to provide overall evidence of epigenetic association with SDOH factors and evaluations for potential heterogeneity in epigenetic association with SDOH factors between the two ancestral groups by using I^2 and P -value (P_{het}) calculated from the Cochran's Q statistic [36]. All analyses were repeated considering an alternative modelling scheme for SDOH factors in which each factor was classified as one of two groups (i.e., a binary variable): high school and above vs. others for educational attainment; none vs. others for personal income; and $>75\%$ vs. others for ADI. We used R package CpGassoc [37] for the linear-regression analysis and $P < 9 \times 10^{-8}$ as the epigenome-wide significance threshold [38]. EWAS results were visualized using circos plots (Circos v.0.69) [39]. Using the EWAS Atlas, we annotated the significant CpGs in our current EWAS [19]. A linear-regression model was fit for the association between the expression level of each gene and the DNAm level of a CpG site while adjusting for age and sex. The Spearman correlations across each paired SDOH factors were illustrated by a heatmap plot. Differentially methylated region (DMR) analysis was performed using the ipDMR method [40]. Potential gene-environment interactions between single-nucleotide polymorphisms (SNPs) (within 1-Mb window) and SDOH factors for DNAm of CpGs were examined using multiple linear regression (DNAm = $\beta_0 + \beta_1 \times \text{SNP} + \beta_2 \times \text{SDOH} + \beta_3 \times \text{SNP} \times \text{SDOH} + \text{other covariates}$) with multiple testing correction by the Benjamini-Hochberg method. All statistical analyses were performed using R.3.6.3 [41] or SAS 9.4 (SAS Institute Inc., Cary, NC) and all statistical tests were two-sided.

Results

Characteristics of the study population

General characteristics, SDOH factors, and clinical features, including primary cancer diagnoses and treatment information, for the study population are provided in Table 1. Study participants included childhood cancer survivors of African ancestry ($n = 258$; median time between cancer diagnosis to blood draw for DNAm = 25.2 years, interquartile range [IQR] = 19.9–32.1 years) and European ancestry ($n = 1,618$; median time between cancer diagnosis to blood draw for DNAm = 27.3 years, IQR = 21.1–33.7 years). For participants of African ancestry, the median age at primary cancer diagnosis was 9.6 (IQR = 4.2–14.4) years and that at blood draw for DNAm was 33.9 (IQR = 29.4–39.6) years; for participants of European ancestry, the ages were 9.0 (IQR = 3.8–14.4) years and 35.3 (IQR = 30.3–42.1) years. The proportion of participants who were female was 53.1% or 40.4% in those of African or European ancestry, respectively. More than 60% of participants were overweight (BMI 25.0 to $<30 \text{ kg/m}^2$, 29.5% in African ancestry and 29.2% in European ancestry) or obese (BMI $\geq 30 \text{ kg/m}^2$, 32.2% in African ancestry and 32.6% in European ancestry). A higher proportion of participants of European ancestry was smokers compared to that of participants of African ancestry (41.8% versus 29.8%, $P < 0.0001$). The three SDOH factors were significantly different between survivors of African ancestry and those of European ancestry, with those of African ancestry having lower educational attainment ($P < 0.0001$), lower personal annual income ($P < 0.0001$), and residing in neighbourhoods with more disadvantaged socioeconomic and physical environment, as measured by ADI ($P < 0.0001$). Furthermore, SDOH factors showed weak correlations with each other ($|\text{Spearman's correlation}| < 0.4$, **Supplementary Figure 3**).

Association of smoking and BMI with SDOH factors among survivors of European ancestry

Smoking behaviour was associated with educational attainment with a significant trend across five categories ($P < 2.2 \times 10^{-16}$). Most (67.9%) survivors with less than a high school education

Table 1. Characteristics of study participants.

	African-ancestry		European-ancestry		<i>P</i> ^a
	(N = 258)		(N = 1,618)		
	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					0.84
<18.5	9	(3.5%)	47	(2.9%)	
18.5- $<$ 25.0	67	(26.0%)	469	(29.0%)	
25.0- $<$ 30.0	76	(29.5%)	472	(29.2%)	
\geq 30.0	83	(32.2%)	527	(32.6%)	
Unknown	23	(8.9%)	103	(6.4%)	
Median BMI (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 blood draw, years					
Median (IQR)	33.9	(29.4-39.6)	35.3	(30.3-42.1)	0.03
Time from diagnosis to blood draw, years					
Median (IQR)	25.2	(19.9-32.1)	27.3	(21.1-33.7)	0.03
Social determinants of health					
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 annual income, US dollars					$<$ 0.0001
None	38	(14.7%)	143	(8.8%)	
$<$ \$20,000	113	(43.8%)	502	(31.0%)	
\$20,000- $<$ 40,000	53	(20.5%)	442	(27.3%)	
\$40,000- $<$ 60,000	19	(7.4%)	256	(15.8%)	
\geq \$60,000	35	(13.6%)	275	(17.0%)	
Area deprivation index, national rank ^b					
Median (IQR)	80.0	(56.0-91.5)	56.0	(36.0-76.0)	$<$ 0.0001
Clinical characteristics					
Primary cancer diagnosis					$<$ 0.0001
Leukaemia	50	(19.4%)	548	(33.9%)	
Lymphoma	57	(22.1%)	639	(39.5%)	
Central nervous system tumour	27	(10.5%)	136	(8.4%)	
Other	124	(48.1%)	295	(18.2%)	
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 site					
Brain	72	(27.9%)	525	(32.4%)	0.15
Chest	81	(31.4%)	513	(31.7%)	0.92
Abdomen/pelvis	76	(29.5%)	391	(24.2%)	0.07

^aChi-square test for categorical variables, Student's *t*-test for continuous variables, and Wilcoxon rank sum test for ranked variables.

^bNational rank in percentiles

were current or former smokers, as compared to 28.9% of college graduates and 14.5% of post-graduates (Table 2). Smoking behaviour was also associated with personal income with ever

smokers ranging between 39.3% and 51.7% and a significant trend across five categories ($P = 8.5 \times 10^{-5}$). Smoking was associated with ADI, and the ever smokers were 29.0%, 39.1%,

Table 2. Association of smoking and BMI with SDOH factors among survivors of European ancestry.

SDOH	Smoking					P for trend ^a
	Total N	Smokers N	(%)	Non-smokers N	(%)	
Educational attainment						<2.2E-16
Below high school	137	93	(67.9)	44	(32.1)	
High school or training after high school	277	136	(49.1)	141	(50.9)	
College	401	200	(49.9)	201	(50.1)	
College graduate	478	138	(28.9)	340	(71.1)	
Post-graduate	179	26	(14.5)	153	(85.5)	
Personal annual income, US dollars						8.50E-05
None	143	74	(51.7)	69	(48.3)	
<\$20,000	502	234	(46.6)	268	(53.4)	
\$20,000-<40,000	442	183	(41.4)	259	(58.6)	
\$40,000-<60,000	256	78	(30.5)	178	(69.5)	
≥\$60,000	275	108	(39.3)	167	(60.7)	
Area deprivation index, national rank in percentiles						4.70E-09
<40	393	114	(29.0)	279	(71.0)	
40–75	611	239	(39.1)	372	(60.9)	
>75	332	167	(50.3)	165	(49.7)	
SDOH BMI						
	Total N	BMI≥25 N	(%)	BMI<25 N	(%)	P for trend ^a
Educational attainment						3.90E-05
Below high school	130	88	(67.7)	42	(32.3)	
High school or training after high school	262	190	(72.5)	72	(27.5)	
College	379	265	(69.9)	114	(30.1)	
College graduate	463	290	(62.6)	173	(37.4)	
Post-graduate	175	91	(52.0)	84	(48.0)	
Personal annual income, US dollars						1.40E-01
None	135	77	(57.0)	58	(43.0)	
<\$20,000	470	317	(67.4)	153	(32.6)	
\$20,000-<40,000	431	275	(63.8)	156	(36.2)	
\$40,000-<60,000	251	177	(70.5)	74	(29.5)	
≥\$60,000	229	153	(66.8)	76	(33.2)	
Area deprivation index, national rank in percentiles						3.80E-06
<40	386	223	(57.8)	163	(42.2)	
40–75	585	394	(67.4)	191	(32.6)	
>75	321	238	(74.1)	83	(25.9)	

Abbreviations: BMI (body mass index), SDOH (social determinant of health)

^aChi-square test for trend

and 50.3% for survivors with low, intermediate, or high SES deprivation, respectively. Similarly, BMI was associated with educational attainment and ADI but not with personal income (Table 2).

Association of DNAm sites with SDOH factors among survivors of European ancestry

The association *p*-values are shown in the circo plot for three series of EWAS analyses among survivors of European ancestry (Figure 1). Quantile-Quantile plots for the distribution of observed and expected *p*-values for EWAS between DNAm CpG sites and SDOH factors showed moderately low genomic inflation factors between 1.20 and 1.35 with adjustments for genetic and methylation principal components, as compared to much higher genomic inflation factors (1.71–3.87) without such adjustments (Supplementary Figure 1). We identified 130

epigenome-wide significant SDOH–CpG associations (Supplementary Table 1), including educational attainment (*n* = 88), personal income (*n* = 23), and ADI (*n* = 19). Thirteen significant DNAm CpGs (cg01731783, cg08840017, cg00385142, cg18754985, cg08064403, cg05659611, cg19859270, cg04180924, cg00010201, cg02978227, cg02657160, cg05575921, cg26768182) were associated with all three SDOH factors and mapped to five genomic regions on chromosomes 2, 3, 5, 9, and 14 harbouring genes, including *GPR55*, *CLDND1*, *CPOX*, *GPR15*, *AHRR*, *PRRC2B*, and *ELMSAN1*, which previously have been reported as associated with smoking exposures (Table 3). After we adjusted for BMI and smoking, all 13 CpGs remained statistically significant (*P* < 0.05) but with attenuated effect sizes for educational attainment (mean effect sizes remaining = 36.8% of the effect sizes unadjusted for BMI and smoking, range = 30.8%–48.8%), personal income (mean = 48.3%,

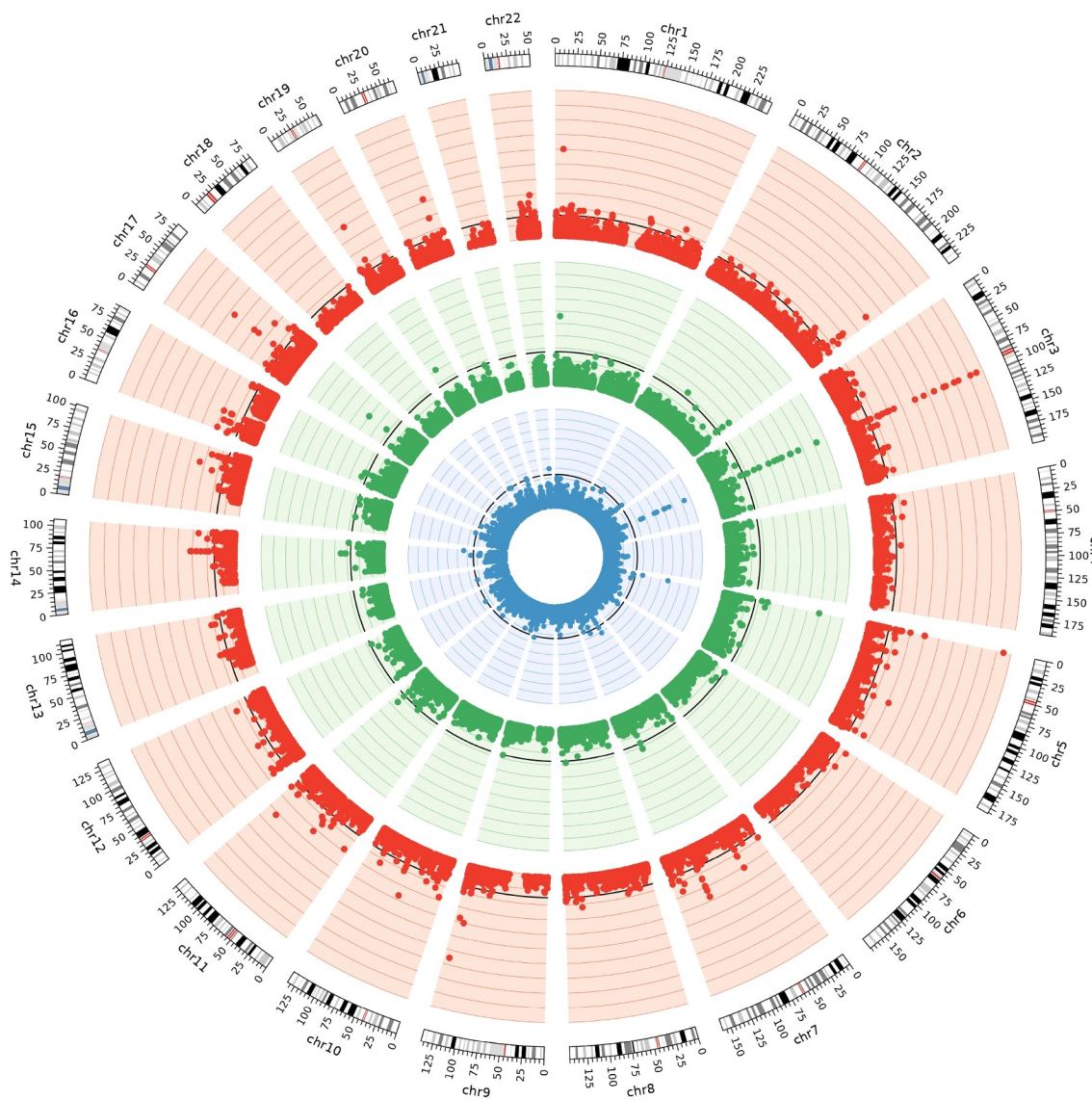


Figure 1. Circos plot for epigenome-wide association studies of SDOH factors among survivors of European ancestry. Outer circle (red): EWAS for educational attainment; middle circle (green): EWAS for personal annual income; inner circle (blue): EWAS for area deprivation index. Each dot depicts $-\log_{10}$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

range = 38.5%-58.8%), and ADI (mean, 43.3%, range = 35.1%-61.5%). In the model for each SDOH factor adjusted for BMI and smoking, there were 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*, $P = 2.0 \times 10^{-8}$) and a significant association between personal income and cg04180924 (chr3, *CPOX*, $P = 5.4 \times 10^{-9}$). No single CpG reached an epigenome-wide significant level for ADI.

Functional implication of SDOH-associated CpG sites

Based on the RNA sequencing data available for 57 samples, we calculated the expression quantitative methylation (eQTM) for each pair of the 13 CpGs and mapped genes (Table 3 and Supplementary Table 2). Three pairs, including cg00385142 (*CLDND1*), cg18754985 (*CLDND1*), and cg05575921 (*EXOC3*), were significant in survivors and were previously known cis-eQTMs (<https://genenetwork.nl/biosqtlbrowser/>). However, the

Table 3. DNA methylation CpG sites that are significantly associated with all three socio-economic factors in survivors of European ancestry.

CpG label	Chr	Position	HGNC gene and functional annotation	CpG Island region	Educational Attainment			Personal Income			Area deprivation index				
					Effect size 1 (SE)	P	Effect size 2 ^a	(% Effect remaining)	Effect size 1 (SE)	P	Effect size 2 ^a	(% Effect remaining)	Effect size 1 (SE)	P	Effect size 2 ^a
cg01731783	chr14	73,745,084	<i>ELMSAM1</i> (5' UTR; 3' UTR)	0.018 (0.002)	1.9E-13	0.006	(30.8%)	0.0133 (0.002)	2.2E-09	0.007	(52.3%)	-0.0007 (0.0001)	3.1E-08	-0.0003	(44.9%)
cg08840017	chr2	230,944,894	<i>GPR55</i> (5' UTR)	0.034 (0.005)	2.9E-12	0.015	(43.9%)	0.0258 (0.004)	7.2E-09	0.014	(55.9%)	-0.0013 (0.0002)	4.0E-08	-0.0008	(58.7%)
cg00385142	chr3	98,517,073	<i>CLDND1</i> (3' UTR; 1st exon)	0.039 (0.004)	1.8E-20	0.019	(48.8%)	0.0246 (0.004)	1.0E-10	0.014	(58.8%)	-0.0015 (0.0002)	7.1E-14	-0.0009	(61.5%)
cg18754985	chr3	98,518,905	<i>CLDND1</i> (3' UTR)	0.054 (0.005)	7.9E-23	0.017	(32.4%)	0.0373 (0.005)	4.6E-14	0.019	(50.8%)	-0.0017 (0.0003)	5.2E-10	-0.0007	(39.1%)
cg08064403	chr3	98,521,413	<i>CLDND1</i> (3' UTR; TSS200; 5' UTR); <i>CPOX</i> (3' UTR); <i>RP11-227H4.5</i> (TSS1500)	0.103 (0.009)	3.4E-28	0.033	(32.2%)	0.0695 (0.008)	2.6E-16	0.034	(49.3%)	-0.0033 (0.0005)	1.3E-12	-0.0014	(41.4%)
cg05659611	chr3	98,521,779	<i>CLDND1</i> (3' UTR; TSS1500; 1st exon; 5' UTR); <i>CPOX</i> (3' UTR); <i>RP11-227H4.5</i> (TSS1500)	0.065 (0.005)	5.0E-31	0.026	(40.6%)	0.0375 (0.005)	1.9E-13	0.019	(50.1%)	-0.0024 (0.0003)	1.3E-17	-0.0012	(51.6%)
cg19859270	chr3	98,532,449	<i>CPOX</i> (3' UTR); <i>GPR15</i> (1st exon)	0.109 (0.008)	4.5E-37	0.037	(34.1%)	0.0671 (0.008)	8.7E-18	0.029	(43.3%)	-0.0032 (0.0004)	4.4E-14	-0.0011	(35.1%)
cg04180924	chr3	98,553,219	<i>CPOX</i> (3' UTR)	0.100 (0.008)	1.9E-35	0.039	(39.0%)	0.0689 (0.007)	3.8E-21	0.037	(53.3%)	-0.0028 (0.0004)	4.4E-12	-0.0011	(38.3%)
cg00010201	chr3	98,568,261	<i>CPOX</i>	0.081 (0.008)	1.4E-25	0.029	(36.1%)	0.0488 (0.007)	2.1E-12	0.022	(44.4%)	-0.0023 (0.0004)	1.3E-09	-0.0007	(31.4%)
cg02978227	chr3	98,573,182	<i>CPOX</i>	0.080 (0.007)	3.5E-32	0.025	(30.8%)	0.0481 (0.006)	8.1E-15	0.020	(40.7%)	-0.0023 (0.0003)	1.8E-12	-0.0008	(35.9%)
cg02657160	chr3	98,592,218	<i>CPOX</i> (3' UTR)	0.055 (0.005)	8.9E-29	0.020	(36.4%)	0.0366 (0.004)	2.0E-16	0.018	(48.3%)	-0.0019 (0.0002)	7.9E-15	-0.0009	(48.9%)
cg05575921	chr5	373,262	<i>AHRR</i> (5' UTR)	0.136 (0.010)	8.1E-37	0.040	(29.4%)	0.0868 (0.010)	1.3E-18	0.036	(42.0%)	-0.0038 (0.0005)	2.5E-13	-0.0014	(35.9%)
cg26768182	chr9	131,397,291	<i>PRRC2B</i> (5' UTR)	0.061 (0.006)	4.6E-26	0.027	(44.0%)	0.0286 (0.005)	8.0E-08	0.011	(38.5%)	-0.0015 (0.0003)	5.1E-08	-0.0006	(40.0%)
Combined							(36.8%)				(48.3%)				(43.3%)

Abbreviations: BMI (body mass index), Chr (chromosome), CpG (5'-cytosine-phosphate-guanine-3'), HGNC (The HUGO genome nomenclature committee), SE (standard error)

^aSmoking/BMI contribution % was calculated by the effect size and additionally adjusted for smoking and BMI divided by the original effect size.

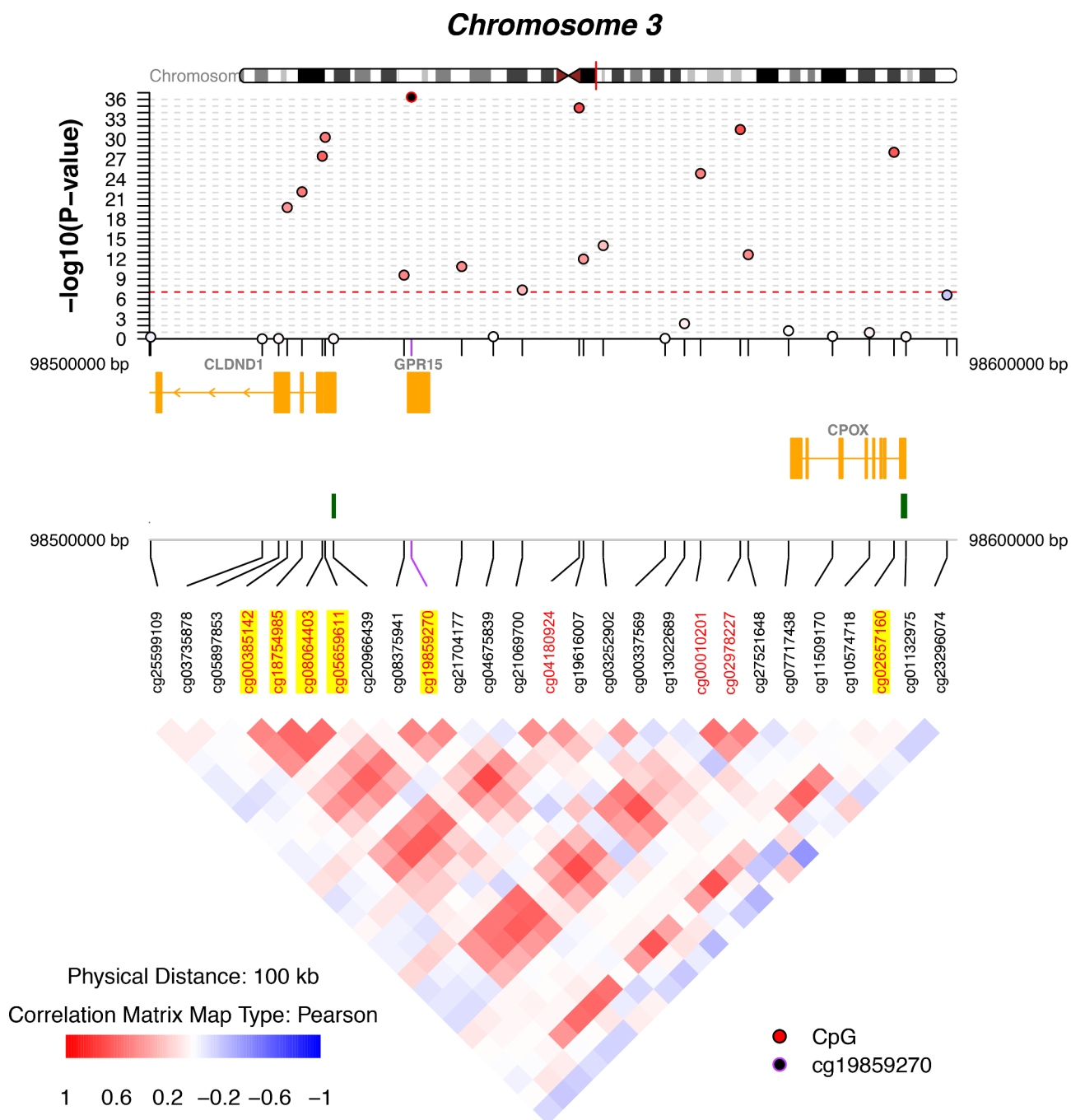


Figure 2. Regional association plots for DNAm CpG sites at chromosome 3 associated with SDOH factors.

Significant epigenome-wide CpGs ($n = 9$) are labelled in red text. Abbreviation: 5'-cytosine-phosphate-guanine-3' (CpG), Social determinants of health (SDOH)

following four known cis-eQTM were not statistically significant in our study: cg19859270 (*CPOX*), cg19859270 (*CLDND1*), cg02657160 (*CLDND1*), and cg02657160 (*CPOX*). Another four pairs [cg08064403 (*CLDND1*), cg05659611 (*CLDND1*), cg19859270 (*GPR15*), and cg00010201 (*CPOX*)] were significant in survivors

but have not been catalogued in the aforementioned biosqtl database.

Regional association with the peak and neighbouring CpGs was illustrated with coMET plot for all five genomic regions (Figure 2 and Supplementary Figure 4). Nine CpGs mapped to the chr3 region (Table 3) and were moderately

correlated with pairwise Pearson correlation coefficients ranging between 0.48 and 0.73. **Figure 2** highlights four CpGs (cg00385142, cg18754985, cg05659611, and cg08064403) that were eQTM for *CLDN1*, one CpG (cg19859270) was an eQTM for *GPR15*, and another (cg02657160) was an eQTM for *CPOX*. We also performed PANTHER pathway analysis but did not find any significant biological pathway with the overrepresentation test based on the group of genes to which all significant CpGs were mapped.

Comparison to previously established findings in the EWAS catalogue

To determine whether the SDOH-associated CpGs identified in the current study have been previously associated with diseases or traits (**Supplementary Table 3**), we cross-referenced the EWAS Atlas. In the EWAS Atlas, a total of 101 associations between blood-based DNAm at 47 CpGs and 24 traits were reported at epigenome-wide significance level ($P < 9 \times 10^{-8}$). Across traits, smoking had the highest number of associated CpGs ($n = 35$), followed by ageing ($n = 8$) and educational attainment ($n = 7$). Among the CpGs associated with SDOH factors in our analysis, cg05575921 was associated with the highest number of different traits ($n = 12$), followed by cg21566642 ($n = 10$) and cg01940273 ($n = 9$). Among the 88 CpGs associated with educational attainment in the current study, 15 overlapped with findings from two previous EWAS studies of educational attainment: detailed association results for 37 CpGs identified in one study and 58 CpGs identified in the other study were provided in the current study (**Supplementary Table 4**).

Association of DNAm sites with SDOH factors among survivors of African ancestry

EWAS analyses were performed in survivors of African ancestry, and no DNAm CpGs reached epigenome-wide significance level ($P < 9 \times 10^{-8}$) due to the limited sample size. However, we assessed the 130 SDOH–CpG associations of genome-wide significance found in survivors of

European ancestry in survivors of African ancestry for comparison (**Supplementary Table 5**). Among the 130 SDOH–CpG associations, 26 were validated in survivors of African ancestry ($P < 0.05$), which included 15 for educational attainments, eight for personal income, and three for ADI.

Secondary analyses including Meta-EWAS, dichotomous SDOH classifications, differentially methylated regions, and gene-SDOH interactions

Meta-analysis of SDOH EWAS among survivors of African ancestry and those of European ancestry showed a total of 448 epigenome-wide significant SDOH–CpG associations (**Supplementary Table 6**); 118 reached the epigenome-wide significance level among survivors of European ancestry alone and were included as one of the 130 SDOH–CpG associations listed in **Supplementary Table 1**. Notably, 88 (19.6%) associations demonstrated at least moderate heterogeneity with $I^2 > 50\%$ between two-ancestral groups. We also considered an alternative modelling scheme for SDOH factors by classifying each factor into two groups (i.e., a binary variable). Meta-EWAS results based on the dichotomous SDOH classifications showed fewer epigenome-wide significant SDOH–CpG associations ($n = 65$) (**Supplementary Table 7**). DMR analysis identified 330, 27, and 32 novel associations with educational attainment, personal income, and ADI, respectively (**Supplementary Table 8**). We examined potential gene-SDOH interactions for DNAm of CpGs within 130 epigenome-wide significant SDOH–CpG associations and identified three statistically significant (FDR < 0.05) interactions between 3 SNPs (rs77289732, rs2470852, and rs2470835) and personal income for DNAm level of cg08064403 and between five SNPs (rs142403317, rs149398072, rs112330421, rs12242855, and rs12255625) and educational attainment for cg22543377, after correction for multiple testing by the Benjamini-Hochberg method (**Supplementary Table 9**).

Discussion

This first social epigenomic study conducted among survivors of childhood cancer examined

the associations between SDOH factors and DNAm. We demonstrated 130 SDOH–CpG associations for educational attainment, personal income, or ADI at epigenome-wide significance. However, we found far fewer CpGs than those previously associated with cancer treatment exposures ($n = 935$) [16,17], and no sites overlapped. The absolute values of effect size for associations between CpGs and SDOH factors (mean = 0.027, SD = 0.023) was significantly smaller than those for associations between CpGs and intensive cancer treatment modalities (mean = 0.11, SD = 0.059). Of the three SDOH factors we examined, educational attainment was the most informative with 88 significant CpG associations.

Most of the SDOH-associated CpGs mapped to genes known to have DNAm associated with smoking exposure, suggesting that social and behavioural exposures to lower educational attainment, lower income, and a higher SES-disadvantaged neighbourhood resemble the effect of tobacco use. Specifically, smoking influences the expression or DNAm of *GPR15* [42], *CPOX* [43], and *CLDND1* [44], suggesting that these gene-environment interactions are implicated along the mechanistic pathway of SDOH factors and health outcomes. This evidence supports our hypothesized framework in which social adversity is more of a distal exposure factor, whereas smoking or any other health behaviour is more of a proximal factor related to health outcomes. Notably, these CpGs remained significant after adjusting for smoking and BMI, though the average effect sizes were smaller than the unadjusted effect sizes for educational attainment, personal income, and ADI, respectively. Moreover, most effect sizes were very small (<5%) with a few up to as high as 13%.

The study's overall findings are consistent with our hypothesis, and the consensus in the field that health behaviours or health behaviour-related factors are the key mediating mechanism between distal social milieu of health and health outcomes [45–47]. Furthermore, some of our robust findings were not related to smoking exposures. For example, cg06359375 was uniquely associated with educational attainment (but no personal income or ADI) at the epigenome-wide significant level, with no attenuation for the effect size after

adjusting for BMI and smoking. This CpG is mapped to the *HPS4* gene, which was previously associated with cognitive function [48].

DNAm associations with educational attainment do not necessarily imply anything about cognitive function but can be proxies for SES, as well as all the associated stressors and comorbidities that accompany lower SES. The effects on cognitive function cannot be disentangled from other social status confounders, such as stress and health behaviours, that are also linked with lower educational attainment.

In non-cancer populations, two large EWAS for SDOH factors have been conducted that specifically addressed educational attainment [8,9]. One large meta-analysis including 27 cohort studies and 10,767 individuals of European ancestry identified 37 CpGs associated with educational attainment in the basic model with nine CpGs remaining in the model adjusted for BMI and smoking status ($P < 9 \times 10^{-8}$). The other meta-analysis with four cohorts and 4,152 individuals from the Netherlands identified 58 CpGs that were associated with educational attainment, and nine CpGs remained significant ($P < 9 \times 10^{-8}$) after adjusting for smoking.

The current study identified 130 CpGs associated with SDOH factors including 88 educational attainment related CpGs, of which only 15 were previously identified as CpGs for educational attainment (**Supplementary Table 10**), suggesting that substantial differences exist in the methylome and its relations with educational attainment between a cancer survivor population and the general population. This was further supported by previously established eQTMs in the general population that were not significant in survivors. To the contrary, we observed strong eQTMs that were not in the biosqtl database.

The observed differences in DNAm associations with educational attainment between cancer survivors and individuals in the general population may have arisen from survivors' experience of cancer and its treatment in childhood. For instance, we speculate that cancer survivors are more resilient to disadvantages associated with lower educational attainment due to having overcome extreme difficulties associated with childhood cancer. Financial hardship associated with childhood cancer may

impede survivors from attaining higher education, and those who achieve higher education under financial hardship may be exceptionally resourceful/successful.

We did not find any significant biological pathways in PANTHER pathways, with the overrepresentation test based on the group of genes to which all significant CpGs were mapped. Biological pathways underpinning the SES influences that increase survivors' vulnerability to adverse health outcomes have not been determined and epidemiological research has thus far yielded only weak and inconsistent evidence on the epigenetics of early-life stress [49]. However, according to a biological embedding of childhood adversity model, childhood stress can be programmed into macrophages through epigenetic markings, post-translational modifications, and tissue remodelling [50]. This could stimulate immune cells to mount an excessive inflammatory response to microbial challenges associated with insensitivity to inhibitory hormonal signals, resulting in a chronic inflammatory state in the body [50]. Moreover, several studies have supported that SES is biologically embedded by showing that low SES across the life course is associated with a blunted pattern of diurnal cortisol production [51], higher level of allostatic load [51], increased inflammatory activity [52–54], and higher pathogen burden [55,56].

Our study has four central limitations. First, the sample size of the survivors of African ancestry was too small to identify ancestry-specific epigenetic associations with SDOH factors, but 20% (26/130) of the epigenome-wide significant CpGs among survivors of European ancestry were validated in the survivors of African ancestry. It is important to note that the observed difference of SDOH-associated CpGs across ancestral groups may be explained by different sociocultural/environmental experiences in addition to genetics. We performed the ancestry-stratified analysis, instead of a pooled analysis, because we found that the DNAm levels for many CpG sites significantly differed between survivors of African ancestry and those of European ancestry. Specifically, 54,125 CpGs were significantly associated with ancestry at the epigenome-wide significance level ($P < 9 \times 10^{-8}$) with a substantial genomic inflation

factor ($\lambda = 5.32$). To supplement our ancestry-stratified analysis, we further conducted a meta-analysis to combine the summary statistics from two stratified results, by following the same analytic strategy used in a recently published multi ethnic EWAS [57].

Second, our analysis was based on a cross-sectional study design, which does not allow assessment of a temporal sequence to establish the causality. Furthermore, we did not have data on the survivors' childhood SDOH factors, such as their parents' educational attainment and income and childhood home addresses to calculate the ADI, all of which potentially affect DNAm throughout their childhood and influence their current (adult) SES. Ideally, DNAm measured at two time points several years apart for the same set of survivors would enable a more rigorous assessment of the effect of baseline SDOH factors (or change of SDOH) on the changes of DNAm between the two time points.

Third, we found fewer CpGs that were significantly associated with income or ADI than were associated with educational attainment. This probably reflects the fact that educational attainment is a relatively stable SDOH factor, compared to income and ADI, which most likely change over the course of survivorship. Therefore, it may be difficult to capture their associations with DNAm in this cross-sectional analysis. In addition, we had significant missingness for the variables of household income and household size in our data set, so we chose personal income as the variable for SDOH modelling. It is important to note that personal income does not always reflect the poverty of the survivor (e.g., a survivor having high household income may have zero or low personal income because he/she does not work for financial, health, or other reasons), and hence the total number of significant epigenetic associations with personal income in our study appeared to be low and should be interpreted with caution.

Fourth, our evaluation of the functional implication of SDOH-associated CpGs was based on a small set of samples that included RNA-sequencing data. We tried to circumvent this limitation by identifying and comparing the eQTM in the public database. Future expanded analysis is needed to confirm or refute the eQTM findings.

In summary, DNAm signatures, many resembling the effect of tobacco use, were associated with SDOH factors among survivors of childhood cancer, suggesting that biologically distal SDOH factors influence health behaviours and modulate the human epigenome. Although disentangling mechanistic pathways remains challenging, future longitudinal studies with SDOH factors and DNAm measured at multiple time points may facilitate the causal inference.

Acknowledgments

The authors acknowledge Dr. Angela McArthur, Department of Scientific Editing, St. Jude Children's Research Hospital for her scientific editing.

Disclosure statement

The authors declare no potential conflicts of interest.

Funding

This work was supported by funding from the V Foundation [Grant # DT2020-014], the National Institutes of Health [Grant # CA021765, CA195547], and the American Lebanese Syrian Associated Charities (ALSAC). The funders of the study had no role in the design or conduct of the study; nor were they not involved in collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; or the decision to submit the manuscript for publication. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Health.

Authorship contributions

Z.W. and I-C.H. designed and performed the research. N.S., J-A.S., Q.D., Z.L., Y.Z., L.H., C-W.H., H.P., C.L.W., K.K.N., K.R.K., D.K.S., Y.Y., M.M.H., L.L.R., I-C.H., and Z.W. collected, analyzed, and interpreted data. N.S., J-A.S., Q.D., Z.L., Y.Z., L.H., C-W.H., H.P., C.L.W., K.K.N., K.R.K., Y.Y., M.M.H., L.L.R., I-C.H., and Z.W. performed the statistical analyses. N.S., J-A.S., I-C.H., and Z.W. wrote the manuscript. M.M.H., L.L.R., I-C.H., and Z.W. contributed administrative, technical, or material support. Z.W. and I-C.H. supervised all aspects of the study.

Statement of prior presentation

Part of this work was preprinted in medRxiv (10.1101/2020.10.30.20223313) and presented in abstract form at the AACR Annual Meeting 2021, which was virtual.

ORCID

Nan Song  <http://orcid.org/0000-0002-9182-1060>

Zhaoming Wang  <http://orcid.org/0000-0001-7556-3869>

References

- [1] Cunliffe VT. The epigenetic impacts of social stress: how does social adversity become biologically embedded? *Epigenomics*. 2016;8:1653–1669.
- [2] Smyth GK. *limma: linear models for microarray data*. In: Gentleman R, Carey VJ, Huber W, et al., editors. *Bioinformatics and computational biology solutions using R and bioconductor*. New York: Springer New York; 2005. p. 397–420.
- [3] Borghol N, Suderman M, McArdle W, et al. Associations with early-life socio-economic position in adult DNA methylation. *Int J Epidemiol*. 2012;41:62–74.
- [4] Lam LL, Emberly E, Fraser HB, et al. Factors underlying variable DNA methylation in a human community cohort. *Proc Natl Acad Sci U S A*. 2012;109(Suppl 2):17253–17260.
- [5] McGuinness D, McGlynn LM, Johnson PC, et al. Socio-economic status is associated with epigenetic differences in the pSoBid cohort. *Int J Epidemiol*. 2012;41:151–160.
- [6] Stringhini S, Polidoro S, Sacerdote C, et al. Life-course socioeconomic status and DNA methylation of genes regulating inflammation. *Int J Epidemiol*. 2015;44:1320–1330.
- [7] Needham BL, Smith JA, Zhao W, et al. Life course socioeconomic status and DNA methylation in genes related to stress reactivity and inflammation: the multi-ethnic study of atherosclerosis. *Epigenetics*. 2015;10:958–969.
- [8] Karlsson Linnér R, Marioni RE, Rietveld CA, et al. An epigenome-wide association study meta-analysis of educational attainment. *Mol Psychiatry*. 2017;22:1680–1690.
- [9] van Dongen J, Bonder MJ, Dekkers KF, et al. DNA methylation signatures of educational attainment. *NPJ Sci Learn*. 2018;3: 7.
- [10] Joehanes R, Just AC, Marioni RE, et al. Epigenetic signatures of cigarette smoking. *Circ Cardiovasc Genet*. 2016;9:436–447.
- [11] Wang X, Zhu H, Snieder H, et al. Obesity related methylation changes in DNA of peripheral blood leukocytes. *BMC Med*. 2010;8:87.
- [12] Dugué PA, Wilson R, Lehne B, et al. Alcohol consumption is associated with widespread changes in blood

- DNA methylation: analysis of cross-sectional and longitudinal data. *Addict Biol.* **2021**;26:e12855.
- [13] Bell JT, Spector TD. DNA methylation studies using twins: what are they telling us? *Genome Biol.* **2012**;13:172.
- [14] Grundberg E, Meduri E, Sandling JK, et al. Global analysis of DNA methylation variation in adipose tissue from twins reveals links to disease-associated variants in distal regulatory elements. *Am J Hum Genet.* **2013**;93:876–890.
- [15] Wang J, Van Den Berg D, Hwang AE, et al. DNA methylation patterns of adult survivors of adolescent/young adult Hodgkin lymphoma compared to their unaffected monozygotic twin. *Leuk Lymphoma.* **2019**;60:1429–1437.
- [16] Song N, Hsu CW, Pan H, et al. Persistent variations of blood DNA methylation associated with treatment exposures and risk for cardiometabolic outcomes in long-term survivors of childhood cancer in the St. Jude Lifetime Cohort. *Genome Med.* **2021**;13:53.
- [17] Song N, Hsu C-W, Pan H, 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* 2020:2020.09.10.20192393.
- [18] Reeves TJ, Mathis TJ, Bauer HE, et al. Racial and ethnic disparities in health outcomes among long-term survivors of childhood cancer: a scoping review. *Front Public Health.* **2021**;9:741334.
- [19] Li M, Zou D, Li Z, et al. EWAS Atlas: a curated knowledgebase of epigenome-wide association studies. *Nucleic Acids Res.* **2019**;47:D983–d8.
- [20] Howell CR, Bjornard KL, and Ness KK, et al. Cohort profile: The St. Jude Lifetime Cohort Study (SJLIFE) for paediatric cancer survivors. *Int J Epidemiol.* **2020**;50(1):39–49.
- [21] Howell CR, Bjornard KL, Ness KK, et al. Cohort profile: The St. Jude Lifetime Cohort Study (SJLIFE) for paediatric cancer survivors. *Int J Epidemiol.* **2021**;50:39–49.
- [22] Kind AJ, Jencks S, Brock J, et al. Neighborhood socioeconomic disadvantage and 30-day rehospitalization: a retrospective cohort study. *Ann Intern Med.* **2014**;161:765–774.
- [23] Kind AJH, Buckingham WR. Making neighborhood-disadvantage metrics accessible - the neighborhood Atlas. *N Engl J Med.* **2018**;378:2456–2458.
- [24] Stolley MR, Restrepo J, Sharp LK. Diet and physical activity in childhood cancer survivors: a review of the literature. *Ann Behav Med.* **2010**;39:232–249.
- [25] Cugnetto ML, Saab PG, Llabre MM, et al. Lifestyle factors, body mass index, and lipid profile in adolescents. *J Pediatr Psychol.* **2008**;33:761–771.
- [26] Qin N, Li Z, and Song N, et al. Epigenetic age acceleration and chronic health conditions among adult survivors of childhood cancer. *J Natl Cancer Inst.* **2020**;113(5):597–605.
- [27] Aryee MJ, Jaffe AE, Corrada-Bravo H, et al. Minfi: a flexible and comprehensive bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics.* **2014**;30:1363–1369.
- [28] Du P, Zhang X, Huang CC, et al. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics.* **2010**;11:587.
- [29] Houseman EA, Accomando WP, Koestler DC, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics.* **2012**;13:86.
- [30] Jaffe AE, Irizarry RA. Accounting for cellular heterogeneity is critical in epigenome-wide association studies. *Genome Biol.* **2014**;15:R31.
- [31] Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics.* **2013**;29:15–21.
- [32] Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics.* **2015**;31:166–169.
- [33] Smyth GK. limma: linear models for microarray data. In: Gentleman R, Carey VJ, and Huber W, et al., editors. *Bioinformatics and computational biology solutions using R and bioconductor. Statistics for biology and health.* New York (NY): Springer; **2005**:229–248. DOI:10.1007/0-387-29362-0_23
- [34] Wang Z, Wilson CL, Easton J, et al. Genetic risk for subsequent neoplasms among long-term survivors of childhood cancer. *J Clin Oncol.* **2018**;36:2078–2087.
- [35] Schisterman EF, Cole SR, Platt RW. Overadjustment bias and unnecessary adjustment in epidemiologic studies. *Epidemiology.* **2009**;20:488–495.
- [36] Higgins J, Thomas J, and Chandler J, et al. *Cochrane handbook for systematic reviews of interventions version 6.2 (updated February 2021).* United Kingdom: Cochrane; **2021**.
- [37] Barfield RT, Kilaru V, Smith AK, et al. CpGassoc: an R function for analysis of DNA methylation microarray data. *Bioinformatics.* **2012**;28:1280–1281.
- [38] Mansell G, Gorrie-Stone TJ, Bao Y, et al. Guidance for DNA methylation studies: statistical insights from the Illumina EPIC array. *BMC Genomics.* **2019**;20:366.
- [39] Krzywinski M, Schein J, Birol I, et al. Circos: an information aesthetic for comparative genomics. *Genome Res.* **2009**;19:1639–1645.
- [40] Xu Z, Xie C, Taylor JA, et al. ipDMR: identification of differentially methylated regions with interval P-values. *Bioinformatics.* **2021**;37:711–713.
- [41] Team. RDC. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; **2010**.
- [42] Kōks S, Kōks G. Activation of GPR15 and its involvement in the biological effects of smoking. *Exp Biol Med (Maywood).* **2017**;242:1207–1212.

- [43] Harlid S, Xu Z, Panduri V, et al. CpG sites associated with cigarette smoking: analysis of epigenome-wide data from the Sister Study. *Environ Health Perspect.* **2014**;122:673–678.
- [44] Maas SCE, Mens MMJ, Kühnel B, et al. Smoking-related changes in DNA methylation and gene expression are associated with cardio-metabolic traits. *Clin Epigenetics.* **2020**;12:157.
- [45] Krieger N. Embodiment: a conceptual glossary for epidemiology. *J Epidemiol Community Health.* **2005**;59:350–355.
- [46] Short SE, Mollborn S. Social determinants and health behaviors: conceptual frames and empirical advances. *Curr Opin Psychol.* **2015**;5:78–84.
- [47] Hertzman C. Commentary on the symposium: biological embedding, life course development, and the emergence of a new science. *Annu Rev Public Health.* **2013**;34:1–5.
- [48] Kuratomi G, Saito A, Ozeki Y, 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.* **2013**;13:276.
- [49] Marzi SJ, Sugden K, Arseneault L, et al. Analysis of DNA methylation in young people: limited evidence for an association between victimization stress and epigenetic variation in blood. *Am J Psychiatry.* **2018**;175:517–529.
- [50] Miller GE, Chen E, Parker KJ. Psychological stress in childhood and susceptibility to the chronic diseases of aging: moving toward a model of behavioral and biological mechanisms. *Psychol Bull.* **2011**;137:959–997.
- [51] Dowd JB, Simanek AM, Aiello AE. Socio-economic status, cortisol and allostatic load: a review of the literature. *Int J Epidemiol.* **2009**;38:1297–1309.
- [52] Ranjit N, Diez-Roux AV, Shea S, et al. Socioeconomic position, race/ethnicity, and inflammation in the multi-ethnic study of atherosclerosis. *Circulation.* **2007**;116:2383–2390.
- [53] Loucks EB, Pilote L, Lynch JW, et al. Life course socioeconomic position is associated with inflammatory markers: the Framingham offspring study. *Soc Sci Med.* **2010**;71:187–195.
- [54] Carroll JE, Cohen S, Marsland AL. Early childhood socioeconomic status is associated with circulating interleukin-6 among mid-life adults. *Brain Behav Immun.* **2011**;25:1468–1474.
- [55] Steptoe A, Shamaei-Tousi A, Gylfe A, et al. Socioeconomic status, pathogen burden and cardiovascular disease risk. *Heart.* **2007**;93:1567–1570.
- [56] Dowd JB, Zajacova A, Aiello A. Early origins of health disparities: burden of infection, health, and socioeconomic status in U.S. children. *Soc Sci Med.* **2009**;68:699–707.
- [57] Jhun MA, Mendelson M, Wilson R, et al. A multi-ethnic epigenome-wide association study of leukocyte DNA methylation and blood lipids. *Nat Commun.* **2021**;12:3987.