DNA methylation as the link between migration and the major noncommunicable diseases: The RODAM study

Article *in* Epigenomics · April 2021

DOI: 10.2217/epi-2020-0329

CITATIONS
3

READS
100

16 authors, including:

Felix P Chilunga
Academisch Medisch Centrum Universiteit van Amsterdam
28 PUBLICATIONS **426** CITATIONS
SEE PROFILE

Peter Henneman
Amsterdam University Medical Centers
196 PUBLICATIONS **8,612** CITATIONS
SEE PROFILE

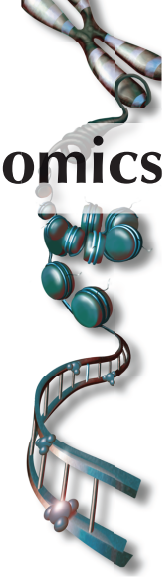
Andrea Venema
Academisch Medisch Centrum Universiteit van Amsterdam
52 PUBLICATIONS **982** CITATIONS
SEE PROFILE

Karlijn A.C. Meeks
National Institutes of Health
106 PUBLICATIONS **1,565** CITATIONS
SEE PROFILE

Research Article

For reprint orders, please contact: reprints@futuremedicine.com

Epigenomics



DNA methylation as the link between migration and the major noncommunicable diseases: the RODAM study

Felix P Chilunga*[,1], Peter Henneman[2], Andrea Venema[2], Karlijn AC Meeks[3], Juan R Gonzalez[4], Carlos Ruiz-Arenas[4], Ana Requena-Méndez[4,5], Erik Beune[1], Joachim Spranger[6], Liam Smeeth[7], Silver Bahendeka[8], Ellis Owusu-Dabo[9], Kerstin Klipstein-Grobusch[10,11], Adebowale Adeyemo[3], Marcel MAM Mannens[2] & Charles Agyemang[1]

[1]Department of Public Health, Amsterdam Public Health Research Institute, Amsterdam University Medical Centers, University of Amsterdam, 1105 AZ Amsterdam, The Netherlands

[2]Department of Clinical Genetics, Amsterdam Reproduction & Development Research Institute, Amsterdam University Medical Centers, University of Amsterdam, 1105 AZ Amsterdam, The Netherlands

[3]Center for Research on Genomics & Global Health, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20894, USA

[4]Barcelona Institute for Global Health (ISGlobal, University of Barcelona), 08003 Barcelona, Spain

[5]Department of Global Public Health, Karolinska Institutet, SE-171 77 Stockholm, Sweden

[6]Department of Endocrinology, Diabetes & Metabolism, Charité-Universitätsmedizin Berlin, 10117 Berlin, Germany

[7]Department of Non-communicable Disease Epidemiology, London School of Hygiene & Tropical Medicine, London, WC1E 7HT, UK

[8]Department of Medicine, MKPGMS-Uganda Martyrs University, 8H33+5M Kampala, Uganda

[9]School of Public Health, Kwame Nkrumah University of Science & Technology, MCFH+R9 Kumasi, Ghana

[10]Julius Global Health, Julius Center for Health Sciences & Primary Care, University Medical Center Utrecht, Utrecht University, 3584 CX Utrecht, The Netherlands

[11]Division of Epidemiology and Biostatistics, School of Public Health, Faculty of Health Sciences, University of The Witwatersrand, 2193 Johannesburg, South Africa

*Author for correspondence: f.p.chilunga@amsterdamumc.nl

Aim: We assessed epigenome-wide DNA methylation (DNAm) differences between migrant and non-migrant Ghanaians. **Materials & methods:** We used the Illumina Infinium® HumanMethylation450 BeadChip to profile DNAm of 712 Ghanaians in whole blood. We used linear models to detect differentially methylated positions (DMPs) associated with migration. We performed multiple post hoc analyses to validate our findings. **Results:** We identified 13 DMPs associated with migration (delta-beta values: 0.2–4.5%). Seven DMPs in *CPLX2*, *EIF4E3*, *MEF2D*, *TLX3*, *ST8SIA1*, *ANG* and *CHRM3* were independent of extrinsic genomic influences in public databases. Two DMPs in *NLRC5* were associated with duration of stay in Europe among migrants. All DMPs were biologically linked to migration-related factors. **Conclusion:** Our findings provide the first insights into DNAm differences between migrants and non-migrants.

First draft submitted: 18 August 2020; Accepted for publication: 29 March 2021; Published online: 23 April 2021

Keywords: DNA methylation • migrants • noncommunicable diseases • RODAM study • sub-Saharan Africans

Migration from low- and middle-income countries (LMIC) to high-income countries (HIC), predominantly for economic reasons, has been increasing in the last few decades with around five million migrants moving to Organization for Economic Co-operation and Development (OECD) countries each year since 2015 [1]. While migration associated with economic growth is likely to raise the average levels of individual well-being, there are concerns about potential deleterious effects on health [2].

Migrants have been shown to be disproportionately affected by cardiometabolic diseases (obesity, diabetes, hypertension, stroke), infectious disease-related cancers (nasopharyngeal, leukemia, cervix) and mental disorders (depression, anxiety disorders) [3–6]. This burden of noncommunicable diseases (NCDs) has been observed to

Future Medicine

increase with duration of stay in host HIC [7]. Adoption of sedentary lifestyle and unhealthy diets, psychosocial stress and limited access to healthcare are thought to be the major driving forces. However, the molecular mechanisms underlying these migration associated changes have not been elucidated [8].

Gene–environmental interactions (through epigenetics) are currently seen as a potential factor mediating the change in lifestyle factors and disease burden but the specific epigenetic changes that occur upon migration are unknown. Epigenetics is the study of heritable yet reversible molecular modifications to DNA expression without altering the DNA sequence [9]. The best understood and most studied epigenetic modification is DNA methylation (DNAm), which is involved in modulation of gene expression via the regulation of transcription factor binding and the attraction of methyl-binding proteins that initiate chromatin compaction and gene silencing [9]. DNAm is known to be affected by the genotype, as well as by environmental factors such as air pollution, tobacco smoke, nutritional factors and nonchemical stressors [10,11]. We, therefore, hypothesize that the vast changes upon migration, including lifestyle factors such as diet and physical activity, and environmental factors such as air quality and reduced presence of pathogenic organisms, alters DNAm patterns among migrants with respect to duration of stay in host countries.

One major problem of studying environmental and lifestyle effects on DNAm is the concurrent influence of the genotype on DNAm. An innovative way of minimizing the concurrent genetic influences on DNAm is by comparing homogenous populations living in LMIC with similar populations that have migrated to HIC (i.e., populations from the same ethnic group or tribe, and same origin in town or village). This innovative design enables identification of migration-specific effects. In addition, sub-Saharan Africans are among the fastest growing group of migrants globally today and are severely impacted by the resulting health transitions, making them a unique population to study effects of migration on DNAm [1,3,8].

We, therefore, assessed for DNAm differences between Ghanaian migrants living in three European countries (Netherlands, Germany and England) and non-migrant Ghanaian counterparts living in rural and urban Ghana. We also further assessed whether DNAm changes among Ghanaian migrants were associated with duration of stay in the three European countries.

Methods

Study population & sample selection

The multicentre Research on Obesity and Diabetes among African Migrants (RODAM) study was initiated in 2012 with the aim of understanding the complex interplay between the environment and genetics in the development of obesity and diabetes among African migrants. The full details of the study have been published elsewhere [12]. In brief, 6385 Ghanaian men and women from the Akan ethnic group were recruited in Europe and in Ghana, of which 5659 met the inclusion criteria (age range 25–70 years, complete questionnaire, physical examination and blood sample profile; Supplementary Figure 1). In Ghana, recruitment of participants in the urban area was conducted in two purposively chosen cities (Kumasi and Obuasi), while recruitment in the rural area was conducted in 15 villages in the Ashanti region. In Europe, participants were recruited from the cities of Amsterdam, Berlin and London. All participants from Europe originated from the same villages (Ashanti region) and towns (Kumasi and Obuasi) in Ghana. The response rates were 53% in Amsterdam, 68% in Berlin, 75% in London, 76% in rural Ghana and 74% in urban Ghana. In this study, we used a subset of the RODAM data for which DNAm data are available (n = 736). These 736 participants were selected for DNA profiling in the initial study based on a case–control design (~300 diabetic cases, ~300 controls, ~135 obese controls). After excluding 24 participants based on quality control criteria, a total of 712 participants remained for current analyses (Supplementary Figure 1). Participants from Europe (Amsterdam, Berlin, London) were categorized as migrants while those from rural and urban Ghana were categorized as non-migrants. All migrants were of first generation. Ethical approval was obtained from ethics committees of involved institutions in Ghana, the Netherlands, Germany and UK before the start of data collection. All participants gave written informed consent.

Phenotypic measurements

A standardized approach for questionnaires, anthropometric measurements and venepuncture samples was used across all study sites. The following measurements were obtained through a structured questionnaire; age, sex, educational attainment, previously diagnosed diabetes mellitus, physical activity levels, alcohol consumption, smoking and duration of stay in Europe. Education was categorized as follows: none or elementary, lower secondary, higher secondary, tertiary. Levels of physical activity were calculated using the Global Physical Activity Questionnaire

(GPAQ), and categorized into low, moderate or high level based on the GPAQ criteria. Alcohol consumption was categorized as any or no consumption. Smoking was categorized into current smokers, past smokers or nonsmokers. Body weight was measured in light clothing and without shoes with SECA 877 scales to the nearest 0.1 kg. Participant's height was measured without shoes with a portable stadiometer (SECA 217) to the nearest 0.1 cm. BMI was calculated as weight (kg) divided by height in m^2. Fasting plasma glucose concentration was measured using an enzymatic method (hexokinase) in mmol/l. Presence of Type 2 diabetes (T2D) was defined using the World Health Organisation (WHO) diagnostic criteria (fasting glucose ≥7.0 mmol/l or current use of medication prescribed to treat diabetes mellitus or self-reported diabetes mellitus). All biochemical analyses were performed in Berlin with an ABX Pentra 400 chemistry analyser (ABX Pentra; Horiba ABX, Germany).

DNAm processing, profiling & quality control

Assessment of the epigenetic profiles, its processing and quality control within the RODAM study were described previously [13,14]. In brief, bisulfite conversion of DNA was conducted with the Zymo EZ DNAm™ kit. The converted DNA was amplified and hybridized on the Infinium® HumanMethylation450 BeadChip, which quantifies DNAm levels of approximately 485,000 CpG sites. Quality control was performed using the *MethylAid* package in R (version 1.4.0). Functional normalization was applied using *minfi* package (version 3.1.0) [15]. Probes annotated to the X and Y chromosomes, known to involve cross-hybridization or to involve common SNPs with a minor allele frequency of ≥0.05 (5%) were removed from the dataset, resulting in a total set of 429,459 CpG sites. Blood cell mixture estimation was based on the method described by Houseman *et al.* [16]. Bioconductor *sva* package was used to construct surrogate variables for removal of unwanted variation [17].

Statistical analysis

Differentially methylated positions

Statistical analysis was carried out using R and Bioconductor packages (package versions are listed in Supplementary file 1) [18]. An epigenome-wide analysis was conducted to assess for differential DNAm between migrants and non-migrants (non-migrants as reference group). Linear regression analyses, with DNAm levels as the dependent variable, were used to identify differentially methylated positions (DMPs) between migrants and non-migrants using *minfi* package. Age, sex, estimated cell types, technical effects (hybridization batch and array position) and surrogate variables were used to adjust our models. We also included BMI and T2D as covariates in the base models to account for previous RODAM reports, which showed an enrichment for obesity and T2D [14,19]. Inflation model fitting was evaluated using QQ-plots (Supplementary Figure 2). False discovery rate (FDR) was used to correct for multiple testing. A 5% FDR was considered statistically significant. DMP analysis was based on M values, while beta values were used for visualization in order to help result interpretation [20].

Differentially methylated regions

DMRcate package was used to detect differentially methylated regions (DMRs). DMRs with ≥ three probes and a 5% FDR were considered as statistically significant.

Post hoc analyses

Based on our study design, it was possible that our results would be confounded by several factors including rural–urban differences within non-migrants, genomic influences from uncharacterized SNPs in Africans (Ghanaians), lack of epigenetic studies on migration to compare our findings with, and varying duration of stay in Europe among migrants. We therefore performed several post hoc analyses to validate our findings.

Rural–urban DNAm differences within non-migrants

Rural and urban Ghanaians were combined into a single group of non-migrants in our study. This would present challenges in interpreting our results since rural and urban environments differ in Ghana, and possible differences between migrants and non-migrants might have more influences from either the rural or urban environment. We therefore performed a separate epigenome-wide analysis to assess for differential DNAm between rural and urban Ghanaians. We applied similar statistical procedures as in the primary analyses.

10.2217/epi-2020-0329

Genomic influences on DNAm differences between migrants & non-migrants

While it is known that the genotype can influence DNAm, genotype data were not available for our study population or Africans at large for further evaluation of our findings. We therefore used *gaphunter* package to test for presence of uncharacterized SNPs in the statistically significant DMPs at a threshold of 0.05 (i.e., 5% difference in consecutive, ordered beta values that defined the presence of a gap signal). . Furthermore, we checked in the GoDMC database (http://mqtldb.godmc.org.uk/) whether statistically significant DMPs were associated with extrinsic SNPs and INDELs (insertion and deletion of nucleotides) in cis or trans of the annotated genes at a Bonferroni multiple test correction threshold of 1.1×10^{-14}.

Replication in independent cohorts

We searched PubMed for cohorts with DNAm data of migration (i.e., data on a homogenous group of migrants in HIC and their non-migrant counterparts in LMIC) to match our study design. However, we were not able to identify any such study. Since we could not verify our findings in independent cohorts, we used the epigenome-wide association studies (EWAS) catalog (http://www.ewascatalog.org/), and genome-wide association studies (GWAS) catalog (https://www.ebi.ac.uk/gwas/), and pathway enrichment in Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases to ascertain whether the genes annotated to statistically significant DMPs were biologically linked to migration-related factors such as environmental factors, lifestyle factors, mental health traits and cardiometabolic traits as presented in our hypothesis [21,22]. Pathway enrichment in GO and KEGG databases was conducted with *MissMethyl* Package (which corrects for probe biases) at 5% FDR.

Correlations between DNAm patterns & gene expression

Study-specific gene expression data were not available for our study population or Africans at large to assess links between the observed DNAm differences and gene expression. Therefore, we assessed gene expression in a publicly available database (The Cancer Genome Atlas; TCGA) using shiny methyl analysis resource tool (SMART) App [23,24]. Correlation coefficients from the TCGA-acute myeloid leukemia dataset (LAML; the only blood borne cancer in the database) were reported.

Duration of stay in Europe among migrants

We used linear regression models to assess whether DNAm in statistically significant DMPs was associated with duration of stay in Europe among migrants. We adjusted for similar covariates as in our primary analyses and used a nominal threshold of $p < 0.05$ to detect statistically significant associations. To minimize potential influences from outlier and influential observations, all observations with a cook's distance $>4/n$ (outliers and influential observations) were excluded from our linear regression models.

Results

Baseline characteristics

A total of 712 individuals were included in the present analyses; 365 were migrants and 347 non-migrants. Mean age of migrants was (50 ± 10) years and non-migrants was (52 ± 10) years ($p < 0.001$). There were fewer female migrants (45.4%) compared with female non-migrants (70.0%; $p < 0.001$). More migrants had attained secondary or university education than non-migrants. The mean length of stay in Europe among migrants was 19.2 ± 10 years. Migrants were more likely to drink alcohol, to smoke and to have obesity compared with non-migrants, while we observed no differences in physical activity between groups (Table 1).

DMPs & regions

The epigenome-wide association study identified 13 DMPs among migrants compared with non-migrants at 5% FDR (Table 2, Figure 1 & Supplementary Figure 3). These DMPs included: cg16411857 in TSS1500 (transcription start site) *of NLRC5*, cg15706807 in TSS200 of *IMPAD1*, cg07839457 in TSS1500 of *NLRC5*, cg07295964 in 5′UTR of *CPLX2*, cg03024619 in first exon of *EIF4E3*, cg13273540 in 5′UTR of *TBL1XR1*, cg16861076 in the body of *BLK*, cg01787285 in the body of *SKI*, cg15723874 in 5′UTR of *MEF2D*, cg01088410 in body of *TLX3*, cg10333808 in 5′UTR of *ST8SIA1*, cg22895601 in TSS1500 of *ANG* and cg13985485 in the body of *CHRM3* as per Illumina platform annotations (IlluminaHumanMethylation450kanno.ilmn12.hg19). The detected differences in DNAm between migrants and non-migrants ranged from 0.2% (cg22895601) to 4.5% (cg15706807, Table 2 & Supplementary Figures 4 & 5).

Table 1. Baseline characteristics of participants.

	All participants (n = 712)	Migrants† (n = 365)	non-migrants‡ (n = 347)	p-value
Demographics, n (%)				
Mean age, SD	51.09 (9.86)	49.89 (9.74)	52.36 (9.84)	<0.001
Sex (Female)	409 (57.36)	166 (45.42)	243 (70.00)	<0.001
Location				
Rural Ghana	104 (14.61)		104 (29.97)	
Urban Ghana	243 (34.13)		243 (70.03)	
Ghanaians in Europe	365 (51.26)	365 (100)		
Education				<0.001
Elementary	226 (32.9)	72 (20.71)	154 (45.56)	
Lower secondary	280 (40.87)	146 (42.08)	134 (39.64)	
Higher secondary	112 (16.35)	75 (21.61)	37 (10.94)	
Tertiary	67 (9.78)	54 (15.56)	13 (3.84)	
Lifestyle factors, n (%)				
Smoking				0.004
Current	15 (2.17)	14 (3.96)	1 (2.96)	
Past	64 (9.26)	31 (8.78)	33 (9.76)	
Alcohol consumption	32 (4.49)	27 (5.12)	5 (1.44)	<0.001
High-level physical activity	285 (44.81)	135 (45.45)	150 (44.25)	0.764
Metabolic factors, n (%)				
Diabetes	256 (39.90)	126 (34.52)	130 (37.46)	0.459
Mean BMI¶	26.09 (5.51)	28.66 (5.58)	25.24 (5.58)	<0.001
Obesity	213 (29.87)	143 (39.12)	70 (20.19)	<0.001
Cells, mean (SD)				
$CD4^+$ T cells	18.13 (5.72)	17.84 (5.47)	18.36 (5.97)	<0.001
$CD8^+$ T cells	10. 92 (4.69)	9.94 (4.53)	11.92 (4.65)	<0.001
Natural killer cells	10.74 (3.56)	10.97 (3.21)	09.21 (3.55)	<0.001
B cells	10. 69 (3.42)	11. 13 (3.55)	10.09 (3.21)	<0.001
Monocytes	8.17 (2.47)	08.20 (0.24)	07.99 (2.50)	<0.001
Granulocytes	45.35 (9.22)	43.04 (8.77)	47.46 (9.16)	<0.001
Length of stay in Europe, years		19.20 (9.64)		

†Migrants = Ghanaians from Amsterdam, Berlin and London.
‡non-migrants = Ghanaians from rural and urban Ghana combined.
¶BMI presented in kg/M^2 as mean with standard deviation.
SD: Standard deviation.

We also identified three DMRs between migrants and non-migrants at 5% FDR. The first DMR was on chromosome 2 between 177.0533 and 177.0545 mb in the body of *HOXD1* consisting of seven probes, which were hypomethylated among migrants (Figure 2 & Supplementary Table 1). The second DMR was on chromosome 7 between 158.938 and 158.939 mb in the TSS of *VIPR2* and its three probes were hypomethylated among migrants (Figure 3 & Supplementary Table 1). The third DMR was on chromosome 22 between 36.0927 and 36.0931 mb in the TSS of *TRIOBP* and contained ten probes, which were less methylated among migrants (Figure 4 & Supplementary Table 1). The detected differences in DNAm within individual CpG sites in these DMRs was similarly smaller and ranged from 0.1% (cg19001226) to 2% (cg05691152, Supplementary Table 1).

Post hoc analyses

Rural–urban DNAm differences within non-migrants

We did not detect any DNAm differences between urban dwelling Ghanaians and rural dwelling Ghanaians at 5% FDR (Supplementary Figures 6–7 & Supplementary Table 2).

10.2217/epi-2020-0329

Table 2. List of differentially methylated positions among migrants compared with non-migrants.

No	CpG ID	Chromosome	Position†	Gene name†	Feature†	Relation to Island†	Delta β value	p-value	FDR
1	cg16411857	16	57023191	*NLRC5*	TSS1500‡	Island	0.021	4.34×10^{-10}	0.000
2	cg15706807	8	58106598	*IMPAD1*	TSS200	Island	-0.045	9.46×10^{-10}	0.000
3	cg07839457	16	57023022	*NLRC5*	TSS1500	N_Shore	0.040	1.29×10^{-8}	0.002
4	cg07295964	5	175223982	*CPLX2*	5′UTR	Island	0.009	7.74×10^{-8}	0.008
5	cg03024619	3	71803308	*GPR27*	1stExon	Island	0.008	3.60×10^{-7}	0.031
6	cg13273540	3	176850227	*TBL1XR1*	5′UTR§	OpenSea	0.016	4.90×10^{-7}	0.031
7	cg16861076	8	11421594	*BLK*	Body	Island	-0.029	5.13×10^{-7}	0.031
8	cg01787285	1	2162682	*SKI*	Body	S_Shore	0.018	6.43×10^{-7}	0.033
9	cg15723874	1	156457945	*MEF2D*	5′UTR	OpenSea	0.013	6.82×10^{-7}	0.033
10	cg01088410	5	170739179	*TLX3*	Body	Island	0.006	8.35×10^{-7}	0.036
11	cg10333808	12	22487459	*ST8SIA1*	5′UTR	Island	0.005	1.26×10^{-6}	0.045
12	cg22895601	14	21131621	*ANG*	TSS1500	OpenSea	0.002	1.31×10^{-6}	0.045
13	cg13985485	1	239550283	*CHRM3*	Body	Island	-0.025	1.35×10^{-6}	0.045

†Annotation were performed via IlluminaHumanMethylation450kanno.ilmn12.hg19. *Homo sapiens* (human) genome assembly GRCh37 (hg19). Hansen KD (2016) IlluminaHumanMethylation450kanno.ilmn12.hg19: Annotation for Illumina's 450k methylation arrays. R package version 0.6.0.
‡TSS1500 (the region from TSS to – 1500 nucleotides upstream of TSS).
§5′UTR (the region of an mRNA that is directly upstream from the initiation codon).
FDR: False discovery rate; TSS: Transcription start site.

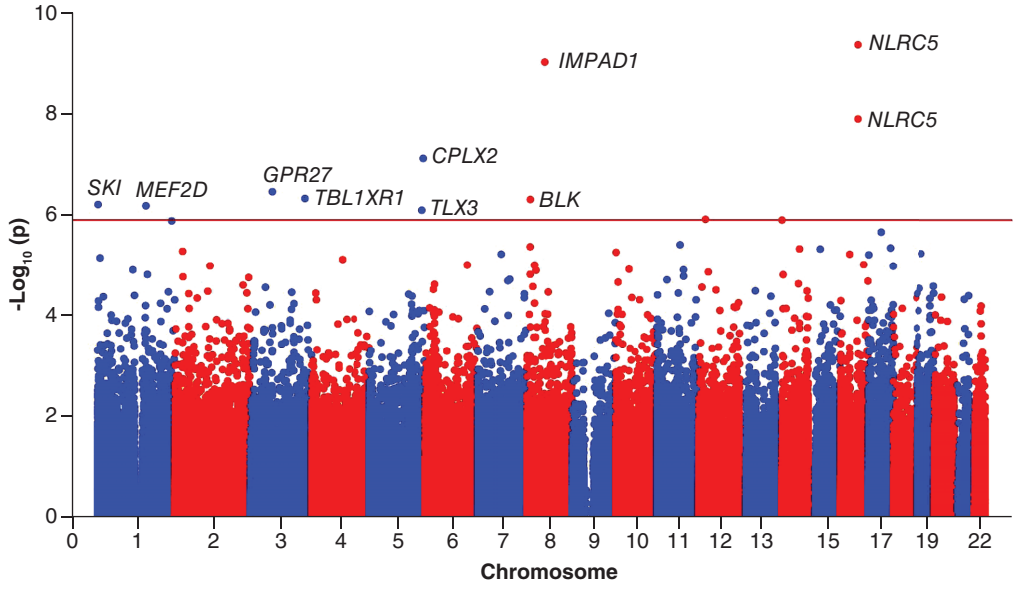


Figure 1. Manhattan plot of differentially methylated positions associated with migration from low- and middle-income countries to high-income countries. All 429,459 CpG sites are presented according to the p-value in the epigenome-wide association study, as well as by chromosomal annotation. Red line is the demarcation line for statistically significant differentially methylated positions at $p < 1.1 \times 10^{-7}$.

Genomic influences on DNAm differences between migrants & non-migrants

Our analysis with *gaphunter* Bioconductor package on the statistically significant DMPs did not return any probes with SNP characteristics at a threshold of 0.05. In the GoDMC database, we found that six DMPs (i.e., cg16411857 in *NLRC5*, cg15706807 in *IMPAD1*, cg07839457 in *NLRC5*, cg13273540 in *TBL1XR1*, cg16861076 in *BLK*, cg01787285 in *SKI*) were associated with extrinsic SNPs and INDELs in cis and trans of their annotated genes at a multiple test correction threshold of 1.1×10^{-14}. The remaining seven DMPs (i.e., cg07295964 in *CPLX2*, cg03024619 in *EIF4E3*, cg15723874 in *MEF2D*, cg01088410 in *TLX3*, cg10333808 in *ST8SIA1*, cg22895601 in *ANG* and cg13985485 in *CHRM3*) did not show such associations (Supplementary Table 3). All three DMRs

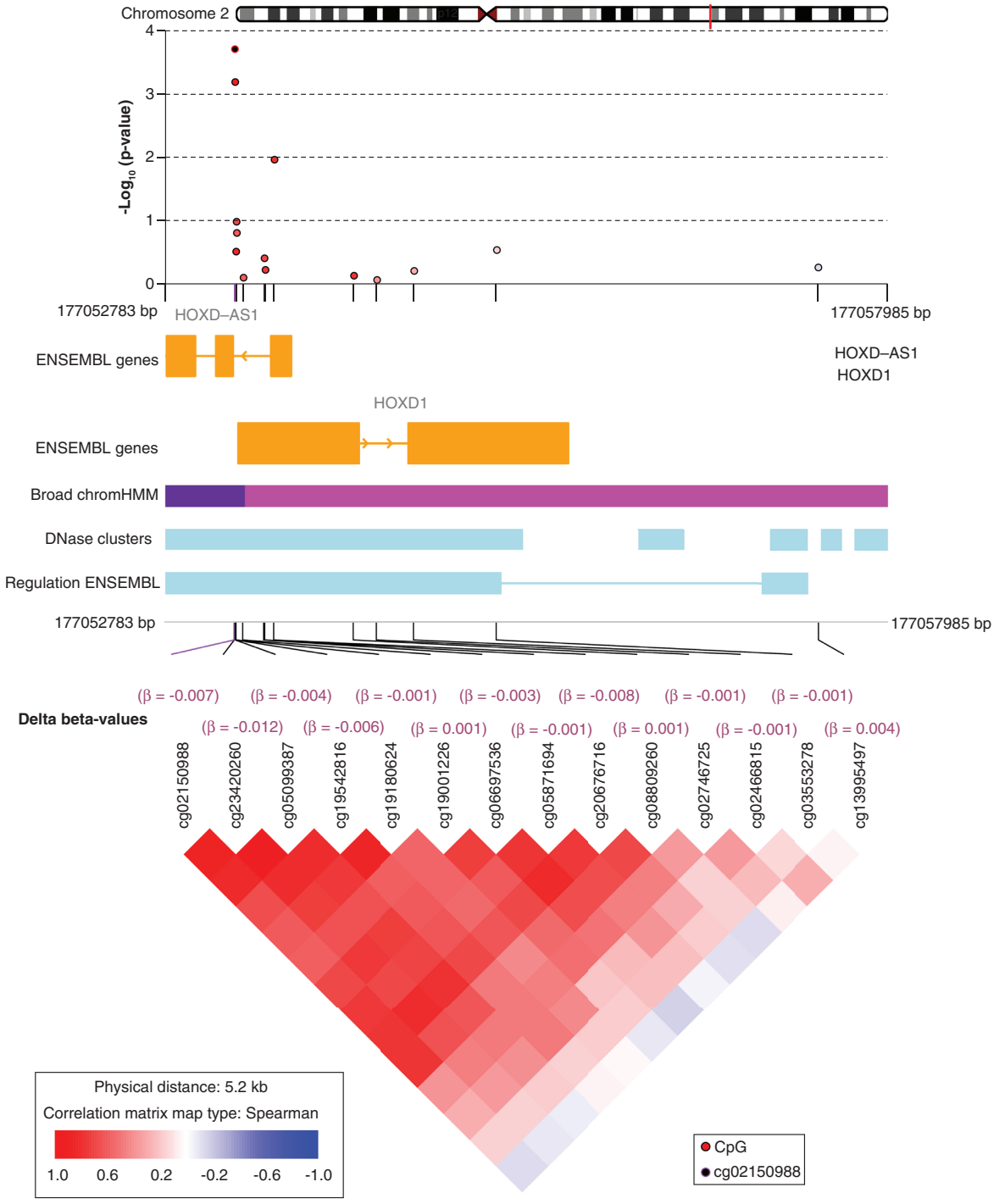


Figure 2. Comet plot of the differentially methylated region identified in *HOXD* gene. Plot shows a differentially methylated region in chromosome 2, between 177.0533 and 177.0545 megabases (mb) obtained via *DMRcate* function in R. Differentially methylated positions within this region are depicted in relation to gene annotation and regulatory elements of those genes. Correlations between differentially methylated positions in this region are also shown.

contained at least two DMPs that were associated with extrinsic SNPs and INDELs in cis and trans of their annotated genes at a multiple test correction threshold of 1.1×10^{-14} (Supplementary Table 4).

10.2217/epi-2020-0329

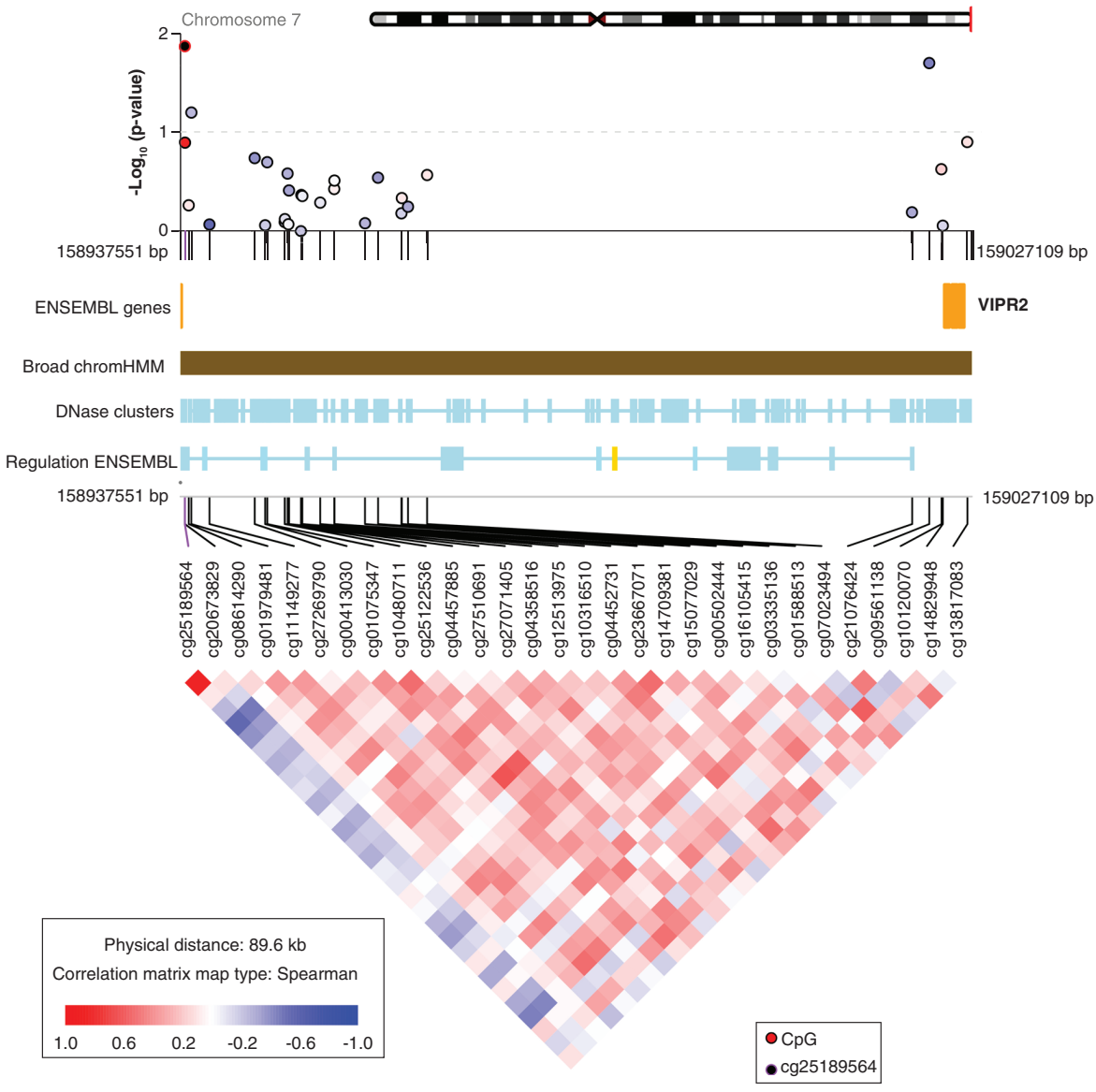


Figure 3. Comet plot of the differentially methylated region identified in *VIPR2* gene. Plot shows a differentially methylated region in chromosome 7, between 158.938 and 158.939 megabases (mb) obtained via *DMRcate* function in R. Differentially methylated positions within this region are depicted in relation to gene annotation and regulatory elements of those genes. Correlations between differentially methylated positions in this region are also shown.

Replication in independent cohorts

Since we could not verify our findings in an independent cohort, we used the EWAS and GWAS catalogs, as well as pathway enrichment in GO and KEGG databases to ascertain whether the genes annotated to the 13 DMPs and three DMRs were linked to migration-related factors such as environmental factors, lifestyle factors, mental health traits and cardiometabolic traits as presented in our hypothesis. We found in the EWAS and GWAS catalogs that genomic/epigenomic variation in the genes annotated to our top DMRs and DMPs were biologically linked to environmental factors (air pollution, nitrogen dioxide exposure, c-reactive protein), lifestyle factors (alcohol consumption, smoking, sedentary behavior, etc.), mental health traits (neuroticism, depression) and cardiometabolic traits (obesity related traits, cholesterol levels, nonalcoholic fatty liver disease, T2D, blood pressure, hepatic growth factor levels and coronary heart disease; Supplementary Tables 5 & 6). Pathway enrichment in KEGG and GO databases did not return any enriched pathways at 5% FDR. However, the pathways that were

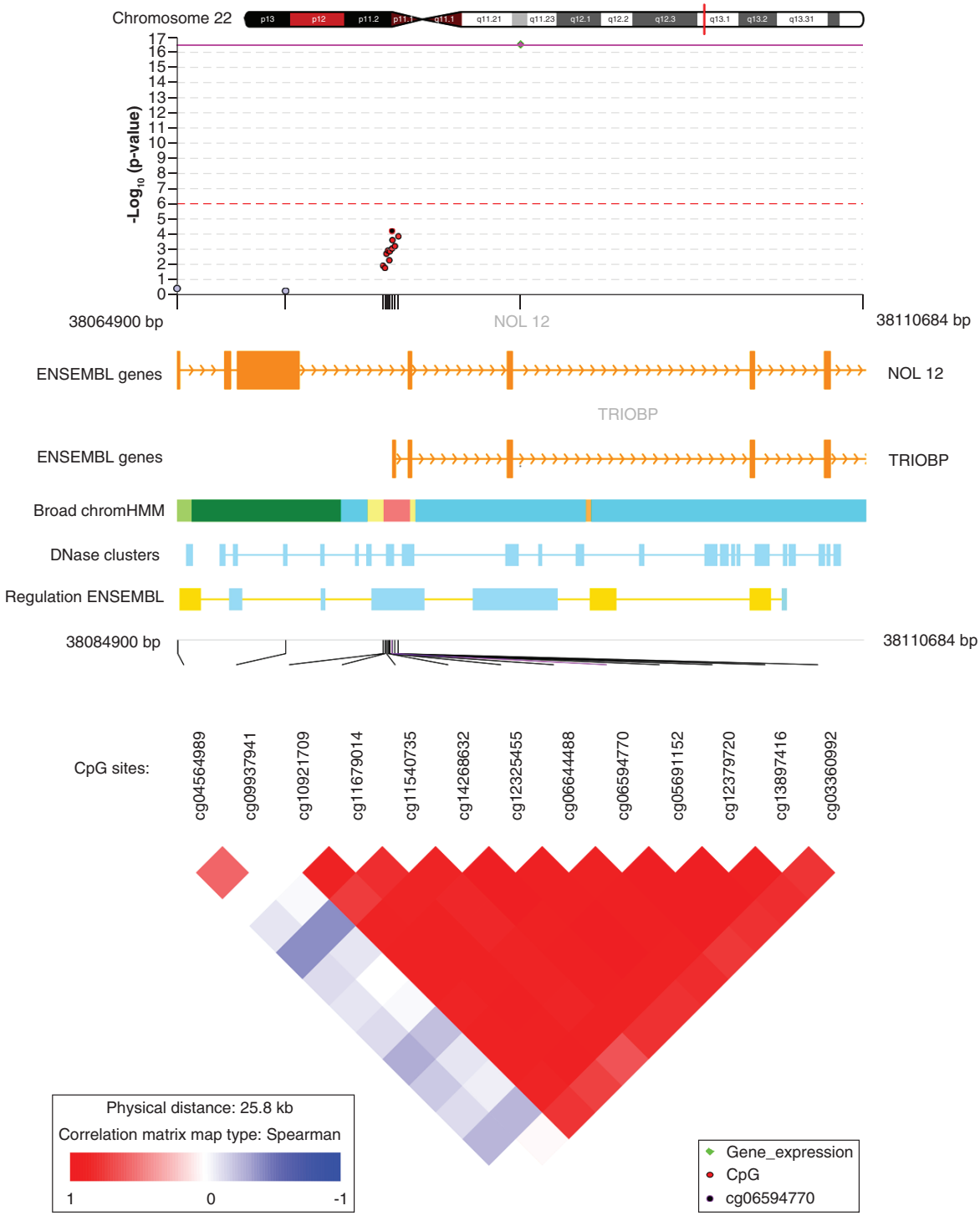


Figure 4. Comet plot of the differentially methylated region identified in *TRIOBP* gene. Plot shows a differentially methylated region in chromosome 22, between 36.0927 and 36.0931 megabases (mb) obtained via *DMRcate* function in R. Differentially methylated positions within this region are depicted in relation to gene annotation and regulatory elements of those genes. Correlations between differentially methylated positions in this region are also shown.

10.2217/epi-2020-0329

returned were in line with our hypothesis and mast cell degranulation (possible allergic reactions to environmental particles), metabolic pathways (glucose and fatty acid synthesis; Supplementary Tables 7 & 8).

Correlations between DNAm patterns

Assessment of the expression of genes annotated to our statistically significant DMPs in the publicly available TCGA-LAML generally followed widely reported trends in the literature, whereby higher DNAm in the promoter region was associated with decreased gene expression, while higher DNAm in the gene body was associated with higher gene expression (Supplementary Figure 8).

Duration of stay in Europe among migrants

Duration of stay in Europe among migrants was positively associated with DNAm in cg16411857 and cg07839457 at nominal threshold of $p < 0.05$. Both DMPs were annotated to the TSS of *NLRC5* (Supplementary Figure 9). The remaining set of the DMPs were not associated with duration of stay in Europe among migrants.

Discussion

Key findings

In our study of epigenome-wide DNAm differences between migrant and non-migrant Ghanaians, we have identified 13 DMPs and three DMRs between migrants and non-migrants, with DNAm differences ranging from 0.1 to 4.5%. Further evaluation of these DMPs and DMRs showed that seven DMPs in *CPLX2*, *EIF4E3*, *MEF2D*, *TLX3*, *ST8SIA1*, *ANG* and *CHRM3* were independent of extrinsic genomic influences in a public database. Additionally, two DMPs in *NLRC5* were associated with duration of stay in Europe among migrants. All DMPs and DMRs were linked to migration-related factors such as air pollution, lifestyle factors, mental health traits and cardiometabolic traits.

Discussion of key findings

In our study of epigenome-wide DNAm differences between migrant and non-migrant Ghanaians, we postulated that vast changes upon migration, including lifestyle factors such as diet and physical activity, environmental factors such as air quality and reduced presence of pathogenic organisms would alter DNAm patterns among migrants as compared with non-migrants. This would help identify molecular (epigenetic) pathways that undergo modification upon migrating from LMIC to HIC.

Of the statistically significant DMPs, seven DMPs in *CPLX2*, *EIF4E3*, *MEF2D*, *TLX3*, *ST8SIA1*, *ANG* and *CHRM3* were independent of extrinsic genomic influences in a publicly available GoDMC database. The gene *CPLX2* encodes for complexin 2, a cytoplasmic protein that regulates the fusion of vesicles to the plasma membrane [25]. The gene *EIF4E3* belongs to a family of translational initiation factors that interact with the 5-prime cap structure of mRNA and recruit mRNA to the ribosome [25]. The gene *MEF2D* is a member of a family myocyte enhancer factor transcription factors with an important role in neuronal and muscle differentiation [25,26]. The gene *TLX3* encodes an orphan homeobox protein that in turn encodes a DNA-binding nuclear transcription factor [25]. The gene *ST8SIA1* encodes a type II membrane protein that catalyzes production of gangliosides, which are important for cell adhesion [25]. The gene *ANG* is a potent inducer of blood vessel formation [25]. The gene *CHRM3* encodes muscarinic receptors. DNAm in these seven genes was generally in line with gene expression trends reported in the literature, whereby higher DNAm in the promoter region was associated with lower gene expression and higher DNAm in the gene body was associated with higher gene expression [27,28]. Previous EWAS and GWAS have shown that these seven genes are linked to migration-related factors [29,30]. For instance, genomic/epigenomic variation in the *CPLX2* gene has been associated with depression onset, and obesity-related traits [29–33], in the *EIF4E3* gene with alcohol consumption, ApoA-I levels and T2D [29,30,34–36], in the *TLX3* gene with smoking, BMI and alcohol consumption [37,38], in the *MEF2D* gene with nitric dioxide pollution, smoking, alcohol consumption, blood pressure and BMI [30,38–41], in the *ST8SIA1* with smoking, nonalcoholic fatty liver disease and coronary artery calcification gene [36,38,42], in the *ANG* gene with nitrogen dioxide pollution, smoking, total lipids and angiogenin levels [40,43,44], and in the *CHRM3* gene with obesity traits, depression, response to stress, hypertension, interleukin levels and alcohol consumption [40,45,46]. All these environmental, lifestyle, mental health and cardiometabolic traits are widely known to differ between migrant and non-migrants or change with migration from LMIC to HIC. For example, tropospheric nitrogen dioxide column densities were shown to be up to five-times greater in western Europe as compared with West African countries (including Ghana) in 2012 [47].

Moreover, migration from LMIC to HIC is known to be accompanied by changes in diet and physical activity, and increased risk of depression and anxiety, as well as increased risk of cardiometabolic diseases [48]. It is therefore possible that these identified pathways are accurate DNAm aberrations associated with migration.

We found that DNAm in two DMPs (cg16411857 and cg07839457) in *NLRC5* was positively associated with duration of stay in Europe among migrants. The gene *NLRC5* plays a role in cytokine response and antiviral immunity through its inhibition of NF-κB activation and negative regulation of type I interferon signaling pathways [25]. Both DMPs were in the promoter of *NLRC5* where higher DNAm was associated with lower gene expression [23,27]. Previous EWAS and GWAS have shown that genomic/epigenomic variation in the *NLRC5* gene was also associated with migration-related factors including alcohol consumption, smoking, tumor necrosis factor and cholesterol levels [33,49–51]. Although cg16411857 and cg07839457 were associated with SNPs and INDELs in the GoDMC database, it is unlikely that these extrinsic genetic associations influenced our findings since the detected effects were in a similar direction (hypermethylation with increased duration of stay), in the same gene (*NLRC5*) and in the same region (transcription start sites; Supplementary Figure 9). It is therefore possible that this finding may accurately reflect on the DNAm changes that occur with respect to duration of stay among migrants.

We discovered that the effect sizes (delta-beta values) were relatively small for the statistically significant DMPs and DMRs. There exist many other genetic and environmental factors outside the studied hypothesis that could drive such small DNAm differences (e.g., seasonal variation in blood cells and age) [16]. Without replication in independent cohorts, the validity of such findings is not fully ascertained. However, we took several steps to make sure that our results were sound. These steps included adjusting for confounders (age, sex, proportion of immune-related cells), paying particular attention to the DMPs that were free of genomic influences (i.e., uncharacterized intrinsic SNPs, as well as extrinsic SNPs and INDELs in cis and trans of the annotated genes), assessing previously reported links to gene expression, as well as assessing links to migration-related factors in previous EWAS and GWAS. It is for this reason that other statistically significant DMPs and DMRs were omitted from our discussion. Our findings therefore provide a foundation for further studies in other migrant populations to confirm our findings, and to further elucidate how migration-specific DNAm differences may influence disease pathology among migrants.

Strengths & limitations

A major strength of this study is the relatively homogenous ethnic SSA Ghanaian study population, which is necessary to study the environmental influences on the epigenome with limited potential bias caused by population stratification. The other strength of our study is that DNAm was assessed in whole blood samples. Given that migration is postulated to affect DNAm across a variety of tissues, our findings therefore act as a pointer to the specific target tissues that need to be studied. The first limitation of our work is that we could neither fully exclude genetic influences on the study nor replicate our findings due to limited genetic and epigenetic data on African populations, hence, there is a possibility of residual confounding. The RODAM study is the first study to assess DNAm in Africans, including migrants. Second, genome-wide gene expression data, which could have enhanced biological interpretation of our results is not available. We tried to utilize gene expression data obtained from normal blood samples in publicly available databases such as the IMETHYL and BIOS QTL browser, however, gene expression data relating to our top probes were not available in the two databases [52,53]. As a result, gene expression data we obtained from the TCGA-LAML (i.e., a cancer dataset) may not fully represent our study tissue. Nevertheless, these publicly available datasets do suggest a link between some of our loci and gene expression. Third, although we removed probes that hybridized to known SNPs, data on SNPs from Africans are limited and some of our results could be due to uncharacterized SNPs in Africans. However, analyses with *gaphunter* package on the top DMPs did not return any probes with SNP-specific patterns [54]. Fourth, the cross-sectional design of our study precludes causal interpretation. Lastly, although we used multiple public databases to assess our findings, these databases are mainly compiled from populations of European descent. There is a possibility that some of the effects reported in these public databases are not present in Africans (Ghanaians) and vice versa.

Conclusion

Our study provides the first insights into DNAm differences between migrants and non-migrants. All observable DNAm differences between migrants and non-migrants were associated with migration-related factors including air pollution, lifestyle factors, mental health traits and cardiometabolic traits. Further studies are needed in other migrant

10.2217/epi-2020-0329

populations (with better analytical techniques such as single cell transcriptomics) to confirm our findings, and to further elucidate how migration-specific DNAm differences may influence disease pathology among migrants.

Future Perspective

With one billion people on the move or having moved in 2018, migration will continue to affect population and individual health well into the future. Since our study has shown that DNAm can be altered with migration, there is an opportunity to develop novel biomarkers of NCD risk, as well as to explore potential drug intervention targets. Further study of the intersection of DNAm, migration and NCDs will lead to discovery of a subset of DNAm aberrations that are most predictive of NCDs upon migration, along with the most promising DNAm targets for NCD therapy.

Summary points

- Migrants face numerous health challenges including a disproportionally high burden of noncommunicable diseases.
- Changes in the environment and in lifestyle factors are suggested to be the major contributing factors.
- DNA methylation could be potential molecular mechanism mediating the change in disease upon migration.
- We identified 13 differentially methylated positions (DMPs) between migrants and non-migrants.
- Seven of these DMPs were independent of extrinsic genomic influences in a public database.
- Two DMPs were positively associated with duration of stay in Europe among migrants.
- All DMPs were biologically linked to migration-related factors including air pollution, lifestyle factors, mental health traits and cardiometabolic traits.
- These are the first insights into DNA methylation aberrations associated with migration.
- Further studies in other migrant ethnic groups are needed to confirm our findings.

Supplementary data

To view the supplementary data that accompany this paper please visit the journal website at: www.futuremedicine.com/doi/suppl/10.2217/epi-2020-0329

Acknowledgments

The authors are very grateful to the Ghanaian volunteers participating in this project.

Financial & competing interest disclosure

This work was supported by the European Commission under the Framework Programme (grant number: 278901) and European Research Council Consolidation (grant number: 772244). FP Chilunga is supported by the Erasmus Mundus Joint Doctorate Programme of the European Union through the Amsterdam Institute of Global Health and Development (AIGHD; grant agreement 2015–1595). A.A. and K.A.C.M. are supported by the Intramural Research Program of the National Institutes of Health in the Center for Research on Genomics and Global Health (CRGGH). The CRGGH is supported by the National Human Genome Research Institute, the National Institute of Diabetes and Digestive and Kidney Diseases, the Center for Information Technology, and the Office of the Director at the National Institutes of Health (1ZIAHG200362). The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Ethical conduct of research

The authors state that they have obtained appropriate institutional review board approval or have followed the principles outlined in the Declaration of Helsinki for all human or animal experimental investigations. In addition, for investigations involving human subjects, informed consent has been obtained from the participants involved.

Data sharing statement

The datasets generated and/or analyzed during the current study are not publicly available due to information that could compromise research participant privacy but are available from the corresponding author on reasonable request.

References

1. Organisation for Economic Co-operation and Development. International Migration Outlook 2019 2019. https://www.oecd-ilibrary.org/content/publication/c3e35eec-en
2. Yusuf S, Reddy S, Ôunpuu S, Anand S. Global burden of cardiovascular diseases: part II: variations in cardiovascular disease by specific ethnic groups and geographic regions and prevention strategies. *Circulation* 104(23), 2855–2864 (2001).
3. Agyemang C, Meeks K, Beune E *et al.* Obesity and type 2 diabetes in sub-Saharan Africans–Is the burden in today's Africa similar to African migrants in Europe? The RODAM study. *BMC Med.* 14(1), 166 (2016).
4. Arnold M, Razum O, Coebergh J-W. Cancer risk diversity in non-western migrants to Europe: an overview of the literature. *Eur. J. Cancer* 46(14), 2647–2659 (2010).
5. Hutchinson G, Haasen C. Migration and schizophrenia: the challenges for European psychiatry and implications for the future. *Soc. Psychiatry Psychiatr. Epidemiol.* 39(5), 350–357 (2004).
6. van der Laan LC, Meeks KA, Chilunga FP *et al.* Epigenome-wide association study for perceived discrimination among sub-Saharan African migrants in europe-the RoDAM study. *Sci. Rep.* 10(1), 1–9 (2020).
7. Boateng D, Agyemang C, Beune E *et al.* Migration and cardiovascular disease risk among Ghanaian populations in Europe: the RODAM study (research on obesity and diabetes among African migrants). *Circ. Cardiovasc. Qual. Outcomes.* 10(11), e004013 (2017).
8. Agyemang C, Addo J, Bhopal R, de Graft Aikins A, Stronks K. Cardiovascular disease, diabetes and established risk factors among populations of sub-Saharan African descent in Europe: a literature review. *Glob. Health.* 5(1), 7 (2009).
9. Barros SP, Offenbacher S. Epigenetics: connecting environment and genotype to phenotype and disease. *J. Dent. Res.* 88(5), 400–408 (2009).
10. Martin EM, Fry RC. Environmental influences on the epigenome: exposure-associated DNA methylation in human populations. *Ann. Rev. Pub. Health.* 39, 309–333 (2018).
11. Daca-Roszak P, Pfeifer A, Żebracka-Gala J *et al.* Impact of SNPs on methylation readouts by Illumina Infinium HumanMethylation450 BeadChip Array: implications for comparative population studies. *BMC Genomics* 16(1), 1003 (2015).
12. Agyemang C, Beune E, Meeks K *et al.* Rationale and cross-sectional study design of the research on obesity and type 2 diabetes among African migrants: the RODAM study. *BMJ Open* 4(3), doi: 10.1136/bmjopen-2014-004877 (2015).
13. Meeks KA, Henneman P, Venema A *et al.* Epigenome-wide association study in whole blood on type 2 diabetes among sub-Saharan African individuals: findings from the RODAM study. *Int. J. Epidemiol.* 48(1), 58–70 (2019).
14. Meeks KA, Henneman P, Venema A *et al.* An epigenome-wide association study in whole blood of measures of adiposity among Ghanaians: the RODAM study. *Clin. Epigenet.* 9(1), 103 (2017).
15. 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* 30(10), 1363–1369 (2014).
16. Houseman EA, Accomando WP, Koestler DC *et al.* DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics* 13(1), 86 (2012).
17. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics.* 28(6), 882–883 (2012).
18. Team RC. R: a language and environment for statistical computing. (2013). https://www.r-project.org/
19. Agyemang C, Beune E, Meeks K *et al.* Rationale and cross-sectional study design of the research on obesity and type 2 diabetes among African migrants: the RODAM study. *BMJ Open.* 4(3), e004877 (2015).
20. Du P, Zhang X, Huang C-C *et al.* Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics* 11(1), 587 (2010).
21. Consortium GO. The Gene Ontology (GO) database and informatics resource. *Nucl. Acid Res.* 32(Suppl. 1), D258–D261 (2004).
22. Kanehisa M (Ed.). The KEGG database. Novartis Foundation Symposium. Wiley Online Library (2002). https://www.genome.jp/kegg/pathway.html
23. Weinstein JN, Collisson EA, Mills GB *et al.* The cancer genome atlas pan-cancer analysis project. *Nat. Genet.* 45(10), 1113–1120 (2013).
24. Li Y, Ge D, Lu C. The SMART App: an interactive web application for comprehensive DNA methylation analysis and visualization. *Epigenetics Chromatin.* 12(1), 1–9 (2019).
25. Safran M, Dalah I, Alexander J *et al.* GeneCards Version 3: the human gene integrator. *Database.* 2010(2010), (2010). https://www.genecards.org/
26. Pan F, Ye Z, Cheng L, Liu JO. Myocyte enhancer factor 2 mediates calcium-dependent transcription of the interleukin-2 gene in T lymphocytes a calcium signaling module that is distinct from but collaborates with the nuclear factor of activated T cells (NFAT). *J. Biol. Chem.* 279(15), 14477–14480 (2004).
27. Lou S, Lee H-M, Qin H *et al.* Whole-genome bisulfite sequencing of multiple individuals reveals complementary roles of promoter and gene body methylation in transcriptional regulation. *Genome Biol.* 15(7), 1–21 (2014).

10.2217/epi-2020-0329

28. Keshet I, Yisraeli J, Cedar H. Effect of regional DNA methylation on gene expression. *Proc. Natl Acad. Sci. USA* 82(9), 2560–2564 (1985).
29. MacArthur J, Bowler E, Cerezo M *et al.* The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucl. Acid Res.* 45(D1), D896–D901 (2017).
30. Battram T, Yousefi P, Crawford G *et al.* The EWAS Catalog: a database of epigenome-wide association studies. (2021). (Epub ahead of print) http://www.ewascatalog.org/
31. Ju C, Fiori LM, Belzeaux R *et al.* Integrated genome-wide methylation and expression analyses reveal functional predictors of response to antidepressants. *Transl. Psychiatry* 9(1), 254 (2019).
32. Comuzzie AG, Cole SA, Laston SL *et al.* Novel genetic loci identified for the pathophysiology of childhood obesity in the Hispanic population. *PLoS ONE* 7(12), e51954 (2012).
33. Hedman ÅK, Mendelson MM, Marioni RE *et al.* Epigenetic patterns in blood associated with lipid traits predict incident coronary heart disease events and are enriched for results from genome-wide association studies. *Circ. Cardiovasc. Genet.* 10(1), e001487 (2017).
34. Richardson TG, Sanderson E, Palmer TM *et al.* Evaluating the relationship between circulating lipoprotein lipids and apolipoproteins with risk of coronary heart disease: a multivariable Mendelian randomisation analysis. *PLoS Med.* 17(3), e1003062 (2020).
35. Chen J, Sun M, Adeyemo A *et al.* Genome-wide association study of type 2 diabetes in Africa. *Diabetologia* 62(7), 1204–1211 (2019).
36. Liu C, Marioni RE, Hedman ÅK *et al.* A DNA methylation biomarker of alcohol consumption. *Mol. Psychiatry* 23(2), 422–433 (2018).
37. Cronjé HT, Elliott HR, Nienaber-Rousseau C, Pieters M. Replication and expansion of epigenome-wide association literature in a black South African population. *Clin. Epigenet.* 12(1), 6 (2020).
38. Joehanes R, Just AC, Marioni RE *et al.* Epigenetic signatures of cigarette smoking. *Circ. Cardiovasc. Genet.* 9(5), 436–447 (2016).
39. 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.* 26(1), e12855 (2021).
40. de F C Lichtenfels AJ, van der Plaat DA, de Jong K *et al.* Long-term air pollution exposure, genome-wide DNA methylation and lung function in the LifeLines cohort study. *Environ. Health Perspect.* 126(2), 027004 (2018).
41. Pulit SL, Stoneman C, Morris AP *et al.* Meta-analysis of genome-wide association studies for body fat distribution in 694 649 individuals of European ancestry. *Hum. Mol. Genet.* 28(1), 166–174 (2019).
42. Wojczynski MK, Li M, Bielak LF *et al.* Genetics of coronary artery calcification among African Americans, a meta-analysis. *BMC Med. Genet.* 14(1), 1–13 (2013).
43. Philibert RA, Beach SR, Lei MK, Brody GH. Changes in DNA methylation at the aryl hydrocarbon receptor repressor may be a new biomarker for smoking. *Clin. Epigenet.* 5(1), 19 (2013).
44. Suhre K, Arnold M, Bhagwat AM *et al.* Connecting genetic risk to disease end points through the human blood plasma proteome. *Nat. Comm.* 8(1), 1–14 (2017).
45. Consortium WTCC. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447(7145), 661 (2007).
46. Eszlari N, Millinghoffer A, Petschner P *et al.* Genome-wide association analysis reveals KCTD12 and miR-383-binding genes in the background of rumination. *Transl. Psychiat.* 9(1), 1–12 (2019).
47. Acker J, Soebiyanto R, Kiang R, Kempler S. Use of the NASA Giovanni data system for geospatial public health research: example of weather-influenza connection. *ISPRS Int. J. Geo-Informat.* 3(4), 1372–1386 (2014).
48. Agyemang C, van den Born B-J. Non-communicable diseases in migrants: an expert review. *J. Travel Med.* 26(2), tay107 (2019).
49. Aslibekyan S, Agha G, Colicino E *et al.* Association of methylation signals with incident coronary heart disease in an epigenome-wide assessment of circulating tumor necrosis factor α. *JAMA Cardiol.* 3(6), 463–472 (2018).
50. Hoffmann TJ, Theusch E, Haldar T *et al.* A large electronic-health-record-based genome-wide study of serum lipids. *Nat. Genet.* 50(3), 401–413 (2018).
51. Chen M-H, Raffield LM, Mousas A *et al.* Trans-ethnic and ancestry-specific blood-cell genetics in 746,667 individuals from 5 global populations. *Cell* 182(5), 1198–213. e14 (2020).
52. Komaki S, Shiwa Y, Furukawa R *et al.* iMETHYL: an integrative database of human DNA methylation, gene expression, and genomic variation. *Human Genome Variat.* 5(1), 1–4 (2018).
53. Bonder MJ, Luijk R, Zhernakova DV *et al.* Disease variants alter transcription factor levels and methylation of their binding sites. *Nat. Genet.* 49(1), 131 (2017).
54. Andrews SV, Ladd-Acosta C, Feinberg AP, Hansen KD, Fallin MD. "Gap hunting" to characterize clustered probe signals in Illumina methylation array data. *Epigenetics & Chromatin.* 9(1), 56 (2016).