Epigenetic signatures of internal migration in Italy.

OXFORD UNIVERSITY PRESS | International Journal of Epidemiology

Epigenetic signatures of internal migration in Italy

Journal:	*International Journal of Epidemiology*
Manuscript ID:	IJE-2014-03-0253.R1
Manuscript Type:	Original Article
Date Submitted by the Author:	09-Aug-2014
Complete List of Authors:	Campanella, Gianluca; Imperial College London, Department of Epidemiology and Biostatistics Polidoro, Silvia; Human Genetics Foundation, Di Gaetano, Cornelia; Human Genetics Foundation, Fiorito, Giovanni; Human Genetics Foundation, Guarrera, Simonetta; Human Genetics Foundation, Krogh, Vittorio; Fondazione IRCCS - Istituto Nazionale dei Tumori, Palli, Domenico; Istituto per lo Studio e la Prevenzione Oncologica (ISPO Toscana), Panico, Salvatore; Università degli Studi di Napoli Federico II, Dipartimento di Medicina Clinica e Sperimentale Sacerdote, Carlotta; Piedmont Reference Centre for Epidemiology and Cancer Prevention (CPO Piemonte), ; Human Genetics Foundation, Tumino, Rosario; Azienda Ospedaliera "Civile - M. P. Arezzo", Registro Tumori ed Unità di Istopatologia Elliott, Paul; Imperial College London, Department of Epidemiology and Biostatistics Matullo, Giuseppe; Human Genetics Foundation, ; Università degli Studi di Torino, Dipartimento di Scienze Mediche Chadeau-Hyam, Marc; Imperial College London, Department of Epidemiology and Biostatistics Vineis, Paolo; Imperial College London, Department of Epidemiology and Biostatistics; Human Genetics Foundation,
Key Words:	Migration, DNA methylation, Developmental origins of disease

Epigenetic signatures of internal migration in Italy

Gianluca Campanella[1,2,†], Silvia Polidoro[3,†], Cornelia Di Gaetano[3], Giovanni Fiorito[3,4], Simonetta Guarrera[3], Vittorio Krogh[5], Domenico Palli[6], Salvatore Panico[7], Carlotta Sacerdote[3,8], Rosario Tumino[9,] Paul Elliott[1,2], Giuseppe Matullo[3,4], Marc Chadeau-Hyam[1,2,§], Paolo Vineis[1,2,3,§]

1. Department of Epidemiology and Biostatistics, Imperial College London, London, UK
2. MRC-PHE Centre for Environment and Health, Imperial College London, London, UK
3. Human Genetics Foundation (HuGeF), Turin, Italy
4. Department of Medical Sciences, University of Turin, Turin, Italy
5. Fondazione IRCCS – Istituto Nazionale dei Tumori, Milan, Italy
6. Istituto per lo Studio e la Prevenzione Oncologica (ISPO Toscana), Florence, Italy
7. Department of Clinical Medicine and Surgery, University of Naples Federico II, Naples, Italy
8. Piedmont Reference Centre for Epidemiology and Cancer Prevention (CPO Piemonte), Turin, Italy
9. Cancer Registry and Histopathology Unit, Azienda Ospedaliera "Civile – M. P. Arezzo", Ragusa, Italy

† Joint first authors
§ Joint last authors

Address for correspondence:

Prof Paolo Vineis

MRC-PHE Centre for Environment and Health

Imperial College London
St Mary's Campus
Norfolk Place, London W2 1PG
United Kingdom

Tel: +44 (0)20 7594 3372

E-mail: p.vineis@imperial.ac.uk

Word count: 166 (abstract) + 2,360 (main text)

Abstract

Background – Observational studies have suggested that the risks of non-communicable diseases in voluntary migrants become similar to those in the host population after one or more generations, supporting the hypothesis that these diseases have a predominantly environmental (rather than inherited) origin. However, no study has been conducted thus far to identify alterations at the molecular level that might mediate these changes in disease risk after migration.

Methods – Using genome-wide DNA methylation profiles from more than 1,000 Italian participants, we conducted an epigenome-wide association studies (EWAS) to identify differences between south-to-north migrants and their origin (southern natives) and host (north-western natives) populations.

Results – We identified several differentially methylated CpG loci, in particular when comparing south-to-north migrants with north-western natives. We hypothesise that these alterations may underlie an adaptive response to exposure differentials that exist between origin and host populations.

Conclusions – Our study is the first large agnostic investigation of DNA methylation changes linked to migratory processes, and shows the potential of EWAS to investigate their biological effects.

Keywords: migration, DNA methylation, developmental origins of disease.

Key messages

1. The risks of many non-communicable diseases in voluntary migrants become similar to those in the host population after one or more generations, but the involvement of alterations at the molecular level (such as DNA methylation) in this process is unclear.
2. Using genome-wide DNA methylation profiles from more than 1,000 Italian participants, we conducted an epigenome-wide association studies (EWAS) to identify differences between southern migrants to north-western Italy, and their origin and host populations.
3. We identified several differentially methylated CpG loci, in particular when comparing south-to-north migrants with north-western natives.
4. We hypothesise that these alterations may be part of an adaptive response to cope with the "mismatch" between early life programming (due to perinatal exposures) and changes that occurred later in life as a result of migration.

Introduction

Observational studies have contributed to consolidate the idea that the risk of many non-communicable diseases in voluntary migrants becomes similar to that in the host (native) population after one or more generations[1-3]. For example, seminal studies have revealed a gradient of increasing incidence of coronary heart disease in Japanese men from Japan to Hawaii to California[3, 4]. These observations have been used to support the hypothesis that these diseases have a predominantly environmental origin (rather than inherited). Nonetheless, differences persist between native and migrant populations. For instance, migrants from non-western countries are more prone to cancers related to infections experienced in early life, and less likely to suffer from cancers commonly associated with a westernised lifestyle[5].

In this paper, we speculate that the observed health differentials might be mediated at the molecular level by changes in DNA methylation. In particular, we hypothesise that these changes are brought about by exposure differentials between the origin and host populations, and that they are instrumental in coping with the "mismatch" between early life programming (due to perinatal exposures) and changes in those same exposures that occurred later in life as a result of migration. This hypothesis is based on the concept that developmental history leaves its mark primarily through potentially reversible epigenetic changes[6]. It is also supported by the observation that disease risks in migrants tend to increase with length of residence in the host population, eventually becoming indistinguishable from those in natives[3]. Among epigenetic changes, DNA methylation is thought to be relatively stable due to its heritability across cell generations, and yet flexible enough to allow differentiation into different cell types, as well as adaptation to stress and the external environment[7-9]. In addition, DNA methylation plays a pivotal role in transcriptional repression and suppression of transcriptional noise[10], and is tightly linked to other epigenetic mechanisms such as histone modifications and chromatin remodelling[11, 12]. DNA methylation levels are associated with environmental and lifestyle exposures such as tobacco smoking[13], and altered DNA methylation patterns have also been implicated in many human diseases[14].

Traditionally, epidemiological studies of migrants endeavoured to elucidate the relative contributions of genetic background, environment, and their interaction[15, 16]. Most studies have focussed on the effects of international migration, since risk factor differentials tend to be larger across countries. In this case, genetic differences between migrants and the host

population may hinder the identification of migration-specific effects. Italy represents an interesting natural experiment, not only for its pronounced economic, social, and environmental south-to-north gradient and the mass migration of labour that took place from the mid-1940s to the 1970s[17, 18], but also for its relative genetic homogeneity (with the possible exception of Sardinia[19]).

Using genome-wide DNA methylation profiles obtained from prospectively collected peripheral blood samples from 1,066 participants in the Italian component of the European Prospective Investigation into Cancer and Nutrition (EPIC-Italy)[20], we present the first epigenome-wide association study (EWAS) to identify DNA methylation changes associated with voluntary south-to-north migration that occurred within Italy in the three decades after the end of the Second World War.

Methods

Study population and sample selection

All participants were recruited between 1993 and 1998 as part of EPIC-Italy[20]. Detailed lifestyle and dietary information was collected at enrolment using self-administered questionnaires and a validated food frequency questionnaire[21], respectively. Anthropometric measurements were obtained at the inclusion visit, as were peripheral blood samples that were aliquoted and stored in liquid nitrogen on the day of collection.

A total of 1,222 genome-wide DNA methylation profiles were acquired as part of three separate prospective case-control studies nested within EPIC-Italy on breast cancer (N = 332), colorectal cancer (N = 338), and myocardial infarction (EPICOR[22], N = 552). Eight profiles were excluded because of unsatisfactory technical quality. A single profile was retained on the basis of technical quality for participants included in more than one study, leaving a total of 1,170 unique profiled participants. Within each study, participants who developed the relevant condition less than one year after blood draw (N = 46), or who developed any kind of haematological malignancy at any time after enrolment (N = 4), were excluded; all remaining subjects were considered healthy at inception. A total of 23 participants were excluded because of incomplete dietary or lifestyle information. To minimise confounding by genetic factors, participants born outside Italy (N = 18) or in the insular region of Sardinia (N = 18) were also excluded. The remaining 1,061 participants were categorised as follows:

1. south-to-north migrants (N = 190), recruited in Turin (N = 148) or Varese (N = 42), and born in any southern Italian region;
2. southern natives (origin population, N = 123), recruited by the two southern Italian EPIC centres of Naples (N = 40) and Ragusa (N = 83), and born in any southern Italian region;
3. north-western natives (host population, N = 543), recruited by the two north-western Italian EPIC centres of Turin (N = 317) and Varese (N = 226), and born in any north-western Italian region.

A total of 205 participants did not fall in any of the above categories, and were excluded from subsequent analyses. Detailed information on the 856 participants included in the study is summarised in Table 1.

Laboratory analyses

Genome-wide DNA methylation analyses using the Illumina Infinium HumanMethylation450 (HM450) platform were carried out at the Human Genetics Foundation (Turin, Italy) according to manufacturers' protocols. Buffy coats stored in liquid nitrogen were thawed, and genomic DNA was extracted using the QIAGEN QIAsymphony DNA Midi Kit. 500 ng of DNA were bisulphite-converted using the Zymo Research EZ-96 DNA Methylation-Gold™ Kit, and hybridised to Illumina Infinium HumanMethylation450 BeadChips. These were subsequently scanned using the Illumina HiScanSQ system, and sample quality was assessed using control probes present on the micro-arrays. Finally, raw intensity data were exported from Illumina GenomeStudio (version 2011.1). Data pre-processing was carried out using in-house software written for the R statistical computing environment. For each sample and each probe, measurements were set to missing if obtained by averaging intensities over less than three beads, or if averaged intensities were below detection thresholds estimated from negative control probes. Background subtraction and dye bias correction (for probes using the Infinium II design) were also performed. The subset of 470,870 probes targeting autosomal CpG loci was selected for further analyses. DNA methylation levels at each locus were expressed as the ratio of intensities arising from methylated cytosines over total intensities.

Statistical analyses

An EWAS was conducted to compare south-to-north migrants to their origin (southern natives) and host (north-western natives) populations, with the objective of characterising

epigenetic adaptation processes subsequent to migration to north-western Italy. For each probe, DNA methylation levels were modelled as dependent variable in a generalised linear model with beta-distributed response using the parameterisation of Ferrari and Cribari-Neto[23]. All models were adjusted for microarray (N = 102) and position on the microarray (N = 12), sex, and case-control status (separately for cancers and myocardial infarction). In place of age at recruitment, models were also adjusted for two continuous variables representing the time to birth and to recruitment of each participant (computed from an arbitrary reference date). Since the difference between these two quantities equals age at recruitment for any choice of reference date, this approach grants an additional degree of freedom to account for possible differences in migration behaviour associated with birth date. The effect of dietary and lifestyle factors, which are radically different in southern regions[24], was investigated using a second set of models additionally adjusted for 25 dietary variables (total energy intake, protein from animal and vegetable sources, fat from animal and vegetable sources, cholesterol, soluble carbohydrates, starch, fibre, alcohol, and vitamins and minerals as listed in Table 1), smoking status, and level of physical activity (categorical variable). To prevent inclusion of highly correlated variables and reduce the number of estimated regression coefficients, dietary variables were subjected to principal component analysis (PCA), and the first 16 principal components (explaining more than 99% of the variance) were included in the models. Multiple comparisons were accounted for by considering a Bonferroni-corrected significance threshold $\alpha = 0.05/470{,}870 \approx 1.1\times10^{-7}$, ensuring a stringent control of the family-wise error rate at level 5%. Candidate CpG loci were additionally filtered as follows. First, probe sequences were aligned to the reference human genome using Bowtie 2[25] to assess the potential to cross-hybridise to multiple genomic locations, thus affecting DNA methylation measurements[26]. CpG loci targeted by cross-hybridising probes (defined as those lacking unique genome alignments, with up to three base mismatches) were excluded from further consideration. Second, potential sources of genetic confounding and context disruption for DNA methylation (such as polymorphisms at the CpG locus) were identified by retrieving known genetic variations and computing the corresponding minor allele frequencies (MAFs) in the European population, based on publicly available data generated by the 1000 Genomes project[27]. As a precautionary measure, CpG loci found within 100 base pairs (bp) of non-rare variants (MAF greater than 1%) were removed from the list of candidates.

Results

As illustrated in Figure 1A, the EWAS identified 20 differentially methylated CpG loci in south-to-north migrants with respect to the origin population (southern natives). Two probe sequences were ambiguously aligned to the reference human genome, and genetic variations were found in the vicinity of nine candidate CpG loci. A total of nine CpG loci were left for further consideration, of which none survived the adjustment for dietary and lifestyle factors (Supplementary Table 1).

Comparison of south-to-north migrants with respect to the host population (north-western natives) revealed 91 differentially methylated CpG loci (Figure 1B). After removal of 23 candidates whose associated probe sequences could not be uniquely aligned to the reference human genome, and of 33 candidates in the proximity of non-rare genetic variations, 35 CpG loci were left for further consideration, and 22 survived the adjustment for dietary and lifestyle factors. Of these, 17 were found to be relatively hypermethylated in south-to-north migrants, and seven were found in the pericentric region on the long arm of chromosome 7 (from the centromere to 6.37×10^7 bp). These loci exhibited a consistent decreasing gradient from south-to-north migrants to southern natives to north-western natives (Figure 2). They were also flanked by several other loci that shared the same direction of association. This region was additionally characterised by PCA of DNA methylation measurements at 43 enclosed CpG loci assayed by the HM450 platform (filtered according to the criteria described above), before and after adjustment for dietary and lifestyle factors. Irrespective of adjustment, the first principal component explained approximately 35% of the variance (Figure 3A), and was the only component explaining more than 10% of the variance. The association of each principal component with migratory status was formally assessed using Kruskal-Wallis rank sum tests. Results were comparable before and after adjustment; however, the second, third, and 36th principal components lost statistical significance after adjustment (Figure 3B). The first principal component was consistently associated with migratory status (p-values 1.71×10^{-8} and 6.28×10^{-6} before and after adjustment, respectively), as was the 15th (p-values 0.043 and 0.041). Scores exhibited a decreasing gradient similar to that observed in Figure 2, albeit less markedly for the 15th principal component (Figure 3C).

Discussion

To our knowledge, this is the first EWAS to examine DNA methylation changes in voluntary migrants. The gamut of alterations observed in south-to-north migrants offers evidence that

important environmental and lifestyle changes may induce molecular adaptation mechanisms to stressors that are inheritable across cell divisions. Some of the differences are evident even after adjustment for dietary and lifestyle factors, suggesting that these DNA methylation changes are not merely ascribable to behaviour modification following migration. Intriguingly, we found DNA methylation changes in south-to-north migrants compared to the host population at several CpG loci located on a large pericentric region on the long arm of chromosome 7. Pericentric regions have long been thought to be transcriptionally inert, but recent evidence suggests that pericentric and centromeric transcripts play an important role in preserving genome stability[28]. Additionally, transcription of pericentric satellites appears to be a general cellular response to external stressors including heat shock, ultraviolet radiation, and oxidative stress[29]. In this light, it appears that molecular consequences of migration may not be limited to specific genes, but may act at a higher complexity level, for example on gene regulatory networks. The gradient observed in Figure 2 may thus epitomise an adaptive mechanism to cope with the "mismatch" between early life programming and exposure changes in later life: before migration, south-to-north migrants and southern natives share common environmental factors that affect DNA methylation patterns and (possibly) differentiate them from northern natives; the amplified response observed after migration might therefore be a consequence of relative abundance or lack of these factors in the host population. Such factors could include, e.g. vitamin D (in relation to more limited sun exposure in northern Italy), other vitamins contained in food, occupational and environmental exposure to pollutants, and even exposure to different infectious agents (with some viruses, for example the hepatitis B virus, being more prevalent in Southern populations). This would not only explain the observed DNA methylation gradient, but it would also be consistent with the "developmental origins of disease" hypothesis[30, 31], and with current understanding of the role of perinatal exposures in health and disease.

The main strengths of our study are its sample size and the relative genetic homogeneity of its participants (all born in Italy), which limits the potential for genetic confounding. Its main limitation is the lack of information regarding the time of migration, from which age at migration and length of stay could be computed and accounted for. Nevertheless, absence of this information is more likely to dilute any observable effect on DNA methylation, rather than leading to false positive results. The biological interpretation of our results could be enhanced were genome-wide gene expression data available for the same subjects. These would allow us to establish whether the observed DNA methylation changes are associated

with gene expression and its regulation, and would thus provide a much deeper understanding of how migration exerts its biological effects at different cellular complexity levels. Despite these limitations, we think this work exemplifies the promising potential of EWAS approaches to elucidate complex and subtle effects of migration at the population level.

Declaration of interests

The authors report no conflicts of interest.

Funding

This work was supported by the 7th European Framework Programme (FP7), grant agreement 308610 (Exposomics) to Paolo Vineis.

EPIC-Italy was financially supported by the Italian Association for Cancer Research (AIRC). Genome-wide DNA methylation profiling of EPIC-Italy samples was financially supported by the Human Genetics Foundation and Compagnia di San Paolo (Turin, Italy).

Paul Elliott is a National Institute for Health Research (NIHR) senior investigator and acknowledges support from the NIHR Biomedical Research Centre at Imperial College Healthcare NHS Trust and Imperial College London. He is supported by the Medical Research Council and Public Health England as part of joint funding for the MRC-PHE Centre for Environment and Health.

Acknowledgments

The authors would like to thank Giuseppe Costa (Epidemiology Unit, Piedmont Region, Turin, Italy) for his insightful comments on a draft version of this paper.

References

1. Geddes M, Balzi D, Buiatti E, Khlat M, Parkin D. Cancer in Italian migrants. *Cancer Causes & Control* 1991; **2**: 133-40.
2. McCredie M. Cancer epidemiology in migrant populations. *Genes and Environment in Cancer*; 1998. p. 298-305.
3. Lassetter JH, Callister LC. The impact of migration on the health of voluntary migrants in western societies: a review of the literature. *Journal of Transcultural Nursing* 2009; **20**: 93-104.
4. Marmot MG, Syme SL, Kagan A, Kato H, Cohen JB, Belsky J. Epidemiologic studies of coronary heart disease and stroke in Japanese men living in Japan, Hawaii and California: prevalence of coronary and hypertensive heart disease and associated risk factors. *American Journal of Epidemiology* 1975; **102**: 514-25.
5. Arnold M, Razum O, Coebergh J-W. Cancer risk diversity in non-western migrants to Europe: an overview of the literature. *European Journal of Cancer* 2010; **46**: 2647-59.
6. Bird A. DNA methylation patterns and epigenetic memory. *Genes & Development* 2002; **16**: 6-21.
7. Reik W. Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature* 2007; **447**: 425-32.
8. Johnstone SE, Baylin SB. Stress and the epigenetic landscape: a link to the pathobiology of human diseases? *Nature Reviews Genetics* 2010; **11**: 806-12.
9. Feil R, Fraga MF. Epigenetics and the environment: emerging patterns and implications. *Nature Reviews Genetics* 2012; **13**: 97-109.
10. Bird AP, Wolffe AP. Methylation-induced repression - belts, braces, and chromatin. *Cell* 1999; **99**: 451-4.
11. Klose RJ, Bird AP. Genomic DNA methylation: the mark and its mediators. *Trends in Biochemical Sciences* 2006; **31**: 89-97.
12. Meissner A, Mikkelsen TS, Gu H, et al. Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* 2008; **454**: 766-70.
13. Zeilinger S, Kuehnel B, Klopp N, et al. Tobacco smoking leads to extensive genome-wide changes in DNA methylation. *PLOS ONE* 2013; **8**: e63812.
14. Egger G, Liang GN, Aparicio A, Jones PA. Epigenetics in human disease and prospects for epigenetic therapy. *Nature* 2004; **429**: 457-63.
15. Haenszel W, Kurihara M. Studies of Japanese migrants. I. Mortality from cancer and other diseases among Japanese in the United States. *Journal of the National Cancer Institute* 1968; **40**: 43-68.
16. Ziegler RG, Hoover RN, Pike MC, et al. Migration patterns and breast cancer risk in Asian-American women. *Journal of the National Cancer Institute* 1993; **85**: 1819-27.
17. Bonifazi C, Heins F. Long-term trends of internal migration in Italy. *International Journal of Population Geography* 2000; **6**: 111-31.
18. Rasulo D, Spadea T, Onorati R, Costa G. The impact of migration in all-cause mortality: the Turin Longitudinal Study, 1971-2005. *Social Science & Medicine* 2012; **74**: 897-906.
19. Francalacci P, Morelli L, Angius A, et al. Low-Pass DNA sequencing of 1200 Sardinians reconstructs European Y-chromosome phylogeny. *Science* 2013; **341**: 565-9.
20. Palli D, Berrino F, Vineis P, et al. A molecular epidemiology project on diet and cancer: the EPIC-Italy prospective study. Design and baseline characteristics of participants. *Tumori* 2003; **89**: 586-93.

21. Pisani P, Faggiano F, Krogh V, Palli D, Vineis P, Berrino F. Relative validity and reproducibility of a food frequency dietary questionnaire for use in the Italian EPIC centres. *International Journal of Epidemiology* 1997; **26**: S152-S60.
22. Fiorito G, Guarrera S, Valle C, et al. B-vitamins intake, DNA-methylation of one carbon metabolism and homocysteine pathway genes and myocardial infarction risk: the EPICOR study. *Nutrition, Metabolism and Cardiovascular Diseases* 2014; **24**: 483-8.
23. Ferrari SLP, Cribari-Neto F. Beta regression for modelling rates and proportions. *Journal of Applied Statistics* 2004; **31**: 799-815.
24. Vineis P, Faggiano F, Riboli E, Berrino F, Pisani P, Crosignani P. Dietary habits, internal migration and social class in a sample of a northern Italian population. *Tumori* 1992; **78**: 235-8.
25. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nature Methods* 2012; **9**: 357-9.
26. Price EM, Cotton A, Lam L, et al. Additional annotation enhances potential for biologically-relevant analysis of the Illumina Infinium HumanMethylation450 BeadChip array. *Epigenetics & Chromatin* 2013; **6**: 4.
27. Altshuler DM, Durbin RM, Abecasis GR, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature* 2012; **491**: 56-65.
28. Hall LE, Mitchell SE, O'Neill RJ. Pericentric and centromeric transcription: a perfect balance required. *Chromosome Research* 2012; **20**: 535-46.
29. Valgardsdottir R, Chiodi I, Giordano M, et al. Transcription of Satellite III non-coding RNAs is a general stress response in human cells. *Nucleic Acids Research* 2008; **36**: 423-34.
30. Waterland RA, Michels KB. Epigenetic epidemiology of the developmental origins hypothesis. *Annual Review of Nutrition* 2007; **27**: 363-88.
31. Kuzawa CW, Quinn EA. Developmental origins of adult function and health: evolutionary hypotheses. *Annual Review of Anthropology* 2009; **38**: 131-47.

Figure legends

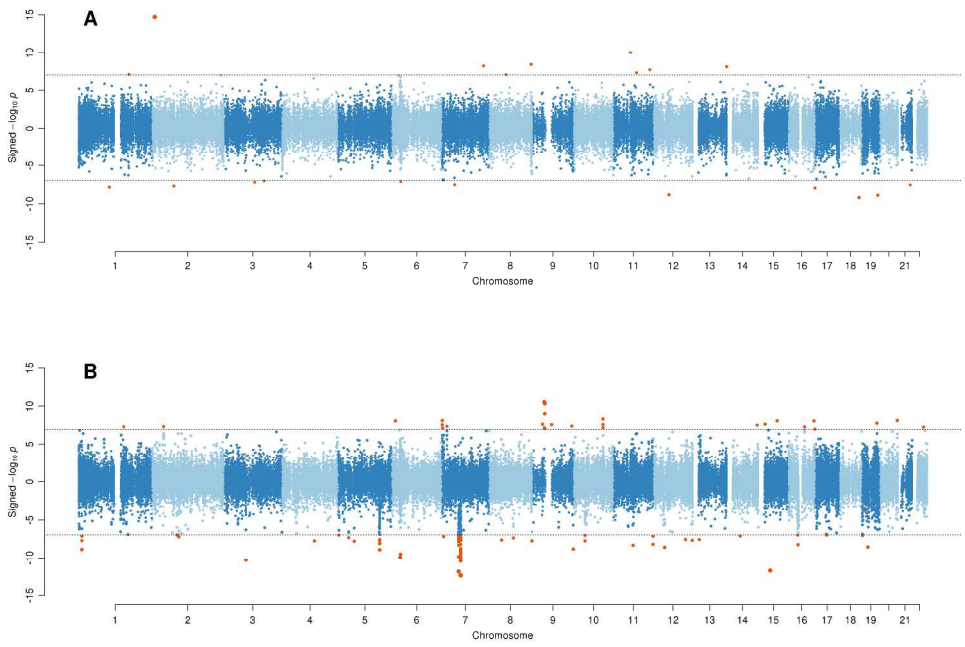
Figure 1. Signed Manhattan plot for the EWAS comparing south-to-north migrants with: (a) the origin population (southern natives); (b) the host population (north-western natives).

Figure 2. Box-and-whisker plot across groups for CpG loci identified by the EWAS in the pericentric region on the long arm of chromosome 7.

Figure 3. PCA of DNA methylation levels assayed in the pericentric region on the long arm of chromosome 7: (a) scree plot; (b) Manhattan plot for the association of PCs with migrant status (Kruskal-Wallis rank sum test); (c) box-and-whisker plot for PCs associated with migrant status ($p < 0.05$) after adjustment for dietary and lifestyle factors.

Table 1. Characteristics of participants included in the study. Counts and percentages are reported for categorical variables, and medians and ranges for continuous variables.

	South-to-north migrants	Southern natives	North-western natives
N	190	123	543
Men	115 (60.5%)	62 (50.4%)	232 (42.7%)
Women	75 (39.5%)	61 (49.6%)	311 (57.3%)
Age (years)	51.2 (35.8 to 66.0)	52.7 (34.7 to 67.9)	54.4 (35.2 to 72.0)
Smoking status			
Never	64 (33.7%)	41 (33.3%)	248 (45.7%)
Current	65 (34.2%)	40 (32.5%)	151 (27.8%)
Former	61 (32.1%)	42 (34.1%)	144 (26.5%)
Physical activity			
1	30 (15.8%)	28 (22.8%)	116 (21.4%)
2	67 (35.3%)	36 (29.3%)	148 (27.3%)
3	67 (35.3%)	51 (41.5%)	251 (46.2%)
4	26 (13.7%)	8 (6.5%)	28 (5.2%)
5	0 (0%)	0 (0%)	0 (0%)
Energy (kcal)	2143.5 (487.7 to 5225.3)	2375.9 (1100.7 to 5476.9)	2184.0 (722.7 to 5309.2)
Protein (g)	85.5 (26.5 to 189.6)	94.1 (39.5 to 190.8)	89.7 (24.4 to 209.3)
From animal sources	53.9 (17.2 to 132.0)	51.3 (12.5 to 103.3)	58.2 (14.3 to 167.8)
From vegetable sources	29.3 (5.6 to 92.0)	40.0 (13.8 to 124.9)	26.3 (5.5 to 70.3)
Fat (g)	77.6 (21.0 to 199.0)	85.9 (36.3 to 178.4)	83.5 (18.9 to 227.7)
From animal sources	41.6 (14.5 to 108.5)	41.9 (12.1 to 105.1)	47.8 (6.4 to 174.6)
From vegetable sources	36.0 (5.6 to 90.5)	42.0 (12.5 to 91.6)	35.0 (4.4 to 91.0)
Cholesterol (mg)	326.7 (77.4 to 772.9)	288.4 (72.9 to 692.5)	358.0 (72.4 to 1002.5)
Available carbohydrates (g)	259.1 (50.8 to 685.9)	320.0 (120.3 to 780.5)	249.2 (48.9 to 760.6)
Soluble carbohydrates	93.6 (15.8 to 284.1)	95.5 (39.0 to 253.0)	98.5 (28.1 to 384.2)
Starch	158.5 (33.7 to 571.6)	209.2 (55.8 to 655.1)	144.0 (5.1 to 460.3)
Fibre (g)	21.0 (3.6 to 51.5)	33.4 (10.5 to 131.2)	19.5 (5.3 to 70.5)
Alcohol (g)	11.5 (0.0 to 90.9)	1.6 (0.0 to 50.4)	11.2 (0.0 to 105.4)
Vitamins			
Vitamin A (μg RE)	944.8 (167.1 to 3467.5)	898.5 (260.1 to 5092.5)	1031.7 (175.4 to 6716.7)
Vitamin B_1 (mg)	0.96 (0.24 to 2.12)	1.22 (0.40 to 4.40)	0.98 (0.33 to 2.18)
Vitamin B_2 (mg)	1.46 (0.30 to 3.12)	1.58 (0.57 to 3.19)	1.58 (0.43 to 4.19)
Vitamin B_3 (mg)	17.8 (5.2 to 39.6)	21.9 (7.8 to 63.9)	17.8 (4.9 to 39.5)
Vitamin B_6 (mg)	1.84 (0.45 to 4.40)	2.28 (0.85 to 5.75)	1.87 (0.46 to 4.12)
Folic acid (μg)	269.8 (48.3 to 775.7)	335.0 (124.4 to 835.0)	261.0 (50.8 to 673.8)
Vitamin C (mg)	127.5 (16.9 to 413.5)	153.2 (63.4 to 672.7)	121.5 (5.9 to 888.5)
Vitamin D (μg)	1.24 (0.13 to 5.85)	1.05 (0.04 to 6.30)	1.34 (0.08 to 12.45)
Vitamin E (mg)	7.65 (1.90 to 24.18)	9.08 (3.78 to 20.04)	7.69 (1.50 to 22.60)
Minerals			
Calcium (mg)	919.9 (279.6 to 2956.3)	834.3 (239.0 to 2050.2)	1020.1 (201.7 to 3833.4)
Iron (mg)	13.8 (2.9 to 35.0)	15.1 (5.8 to 42.9)	13.8 (4.8 to 31.9)
Phosphorus (mg)	1342.6 (382.9 to 2945.3)	1558.9 (576.4 to 3816.1)	1413.4 (487.0 to 3393.8)
Potassium (mg)	3207.2 (690.1 to 7023.0)	3564.4 (1414.5 to 8445.0)	3271.4 (982.1 to 8882.6)
Sodium (mg)	2164.9 (359.2 to 9837.6)	2365.8 (729.8 to 8879.0)	2227.3 (728.5 to 8636.5)
Zinc (mg)	11.7 (4.0 to 26.1)	14.2 (4.4 to 42.9)	12.4 (3.6 to 32.8)



Signed Manhattan plot for the EWAS comparing south-to-north migrants with: (a) the origin population (southern natives); (b) the host population (north-western natives).
419x297mm (300 x 300 DPI)