Gene methylation parallelisms between peripheral blood cells and oral mucosa samples in relation to overweight.

Rodrigo San-Cristobal ¹, Santiago Navas-Carretero ^{1,2}, Fermín I. Milagro ^{1,2,3}, J. Ignacio Riezu-Boj ^{1, 3}, Elizabeth Guruceaga ^{3, 4}, Carlos Celis-Morales ⁵, Katherine M Livingstone ⁵, Lorraine Brennan ⁶, Julie A. Lovegrove ⁷, Hannelore Daniel ⁸, Wim H. Saris ⁹, Iwonna Traczyk ¹⁰, Yannis Manios ¹¹, Eileen R. Gibney ⁶, Michael J. Gibney ⁶, John C. Mathers ⁵, J. Alfredo Martinez ^{1,2,3}

¹ Department of Nutrition, Food Science and Physiology. Centre for Nutrition Research, University of Navarra, Pamplona, Spain.

² CIBER Fisiopatología Obesidad y Nutrición (CIBERObn), Instituto de Salud Carlos III, Madrid, Spain

³ Instituto de Investigación Sanitaria de Navarra (IdiSNA), Pamplona, Spain.

⁴ Unit of Proteomics, Genomics and Bioinformatics, Center for Applied Medical Research (CIMA), University of Navarra, Pamplona, Spain.

⁵ Human Nutrition Research Centre, Institute of Cellular Medicine, Newcastle University, Campus for Ageing and Vitality, Newcastle Upon Tyne, UK

⁶ UCD Institute of Food and Health, University College Dublin, Belfield, Dublin 4, Republic of Ireland

⁷ Hugh Sinclair Unit of Human Nutrition and Institute for Cardiovascular and Metabolic Research, University of Reading, Reading, UK

⁸ ZIEL Research Center of Nutrition and Food Sciences, Biochemistry Unit, Technische Universität München, Germany

- 22 9 Department of Human Biology, NUTRIM, School for Nutrition and Translational
23 Research in Metabolism, Maastricht University Medical Centre, Maastricht, The Netherlands
- 24 10 Department of Human Nutrition, Faculty on Health Sciences, Medical University of
25 Warsaw, Poland.
- 26 11 Department of Nutrition and Dietetics, Harokopio University, Athens, Greece.

Abstract

Epigenetics has an important role in the regulation of metabolic adaptation to environmental modifications. In this sense, the determination of epigenetic changes in non-invasive samples during the development of metabolic diseases could play an important role in the procedures in primary healthcare practice. To help translate the knowledge of epigenetics to public health practice the present study aims to explore the parallelism of methylation levels between white blood cells and buccal samples in relation to obesity and associated disorders. Blood and buccal swaps samples were collected from a subsample of the Spanish cohort of the Food4Me study. Infinium HumanMethylation450 DNA Analysis was carried out for the determination of methylation levels. Standard deviation for β -values method and concordance correlation analysis were used to select those CpG which showed best parallelism between samples. A total of 277 CpGs met the criteria and were selected for an enrichment analysis and a correlation analysis with anthropometrical and clinical parameters. From those selected CpGs, 4 presented high associations with: BMI (cg01055691 in *GAP43*; $r=-0.92$ & $\rho=-0.84$ for blood; $r=-0.89$ & $\rho=-0.83$ for buccal sample); HOMA-IR (cg00095677 in *ATP2A3*; $r=0.82$ & $\rho=-0.84$ for blood; $r=-0.8$ & $\rho=-0.83$ for buccal sample); and leptin (cg14464133 in *ADARB2*; $r=-0.9182$ & $\rho=-0.94$ for blood; $r=-0.893$ & $\rho=-0.79$ for buccal sample). These findings demonstrate the potential application of non-invasive buccal samples in the identification of surrogate epigenetic biomarkers, and identify methylation sites in *GAP43*, *ATP2A3* and *ADARB2* genes as potential targets in relation to overweight management and insulin sensibility.

Keywords: Methylation, obesity, parallelism, mirror tissues, surrogate tissues

Introduction

Obesity prevalence rates increase each year worldwide, resulting in a concomitant rise of associated metabolic diseases [13]. There are many possible causes, including the westernization of the dietary patterns, characterized by an increment of energy-dense food

consumption, as well as the wider adoption of sedentary lifestyles, which induce a chronic energy misbalance [48]. Epigenetics appear to be an important dynamic element in the adaptation to environmental changes and in the triggering of specific physiological responses [9].

Several studies have shown that the epigenome has the potential to be modified by exposure to a range of nutritional factors [34]. A study by van Dijk et al, [15] has demonstrated that these epigenetic modifications are associated with a increase risk of developing obesity, as well as other non-communicable diseases such as type 2 diabetes, hypertension, dyslipidaemias or cardiovascular events [12, 16, 35, 43]. Indeed, epigenetic markers, and more specifically DNA methylation levels, may represent a potential marker for disease, which have been studied to identify a predictive role of methylation changes in pathways related to metabolic diseases [34]. Because DNA methylation patterns can be cell-type specific, these studies have investigated changes within specific cell types and tissues, such as white adipose tissue, muscle, pancreas and blood [1, 3, 14, 38, 44, 45, 54], where the functional effect is assumed to occur [32]. However, in some cases the sampling procedures are invasive and it is not possible to carry out these procedures in primary healthcare practice. For these reasons, and in order to translate the knowledge about epigenetics to public health practice and policies, there is the need to evaluate other tissues [5]. This strategy would facilitate the use of alternative and more feasible techniques from non-invasive tissues, which may “mirror” pattern of epigenetic markers in other specific tissues of interest (i.e., skeletal muscle, liver, adipose tissue, etc). To achieve this goal, research has evaluated the epigenetic modifications of obesity in less invasive and more commonly collected tissues, such as peripheral blood mononuclear cells [8, 20, 33, 53, 56]. Most recently, other studies have analysed DNA methylation patterns in peripheral tissues, blood and subcutaneous adipose tissue, to explore the parallelism in the development of different metabolic disorders related to obesity [24, 45]. Indeed, epigenetic markers in DNA from buccal swabs have been used to characterize novel

biomarkers in non-invasive, easily accessible samples. Moreover, buccal samples have been suggested as better informative surrogates than blood for epigenome-wide association studies [27].

The development of high-throughput technologies has also facilitated the analysis of different tissues to compare changes at an epigenomic level, which may have been produced by, or are associated with, metabolic diseases. These methodologies could be used to develop a “personalised nutrition” approach to provide a tailored assessment for the prevention of non-communicable diseases, as well as for designing new treatments [6].

Taking into account these premises, we propose to examine the eventual parallelism in DNA methylation levels between white blood cells (WBC) and buccal samples, at specific CpGs concerning obesity and associated disorders.

Material and methods

Study subjects

A subset of volunteers were selected from the Spanish cohort of the Food4Me study, which is a randomized, longitudinal, multicentre trial of personalised nutrition advice, delivered through online assessment with four levels of intervention and two intensities [10]. A total of 43 volunteers were invited to attend the University of Navarra to validate their self-reported measures carried out during the main study. Anthropometric measurements were taken by trained researchers. Blood samples were also collected after a 10-12 hours overnight fast to perform the biochemical determinations and validate the genetic data. From those 43 volunteers a subsample of 10 volunteers were selected to take part in the present study. For this purpose, buccal swaps samples were collected to isolate DNA and to determine DNA methylation levels.

Procedures

Anthropometric measurements: Height (cm) was measured with a calibrated stadiometer (SECA 220 Hamburg, Germany) to the nearest 0.1 cm. Body weight (kg) and body composition were assessed by Bioelectric Impedance Analysis (Tanita BC 418, TANITA Corp, Tokyo, Japan) to the nearest 0.1 kg for both. Anthropometric measurements of waist, hip and upper leg circumference (cm) were also recorded by trained staff with a commercial inelastic and flexible tape to the nearest 0.1 cm. Global adiposity was estimated by body mass index (BMI; kg/m²) and percentage of fat mass. Waist-hip ratio was calculated to assess central adiposity.

Biochemical analyses: Plasma, serum and WBC were separated from whole blood samples by centrifugation, 10 minutes at 1500g at room temperature, and immediately frozen at -80°C until assay. Serum concentrations of glucose, total and HDL cholesterol, triglycerides, alanine aminotransferase (ALT), aspartate aminotransferase (AST), and homocysteine were measured by standardized methods in a Pentra C-200 autoanalyser (HORIBA ABX, Madrid, Spain): Insulin, C reactive protein (CRP), cortisol and leptin concentrations were determined by enzyme-linked immunosorbent assay (ELISA) in a Triturus autoanalyser (Grifols SA, Barcelona, Spain). Low-density lipoprotein (LDL) cholesterol and Homeostasis Model Assessment Index (HOMA-IR) data were calculated according to previous described formulae [19, 29].

DNA Isolation from different tissues: DNA isolation from WBC was performed by using MasterPure DNA Purification Kit for Blood Version II (Epicenter, Madison, Wis., USA) following the manufacturer's protocols. DNA from buccal swaps was isolated with Maxwell 16 Buccal Swab LEV DNAPurification Kit (Promega, Madison, WI, USA) following the manufacturer's instructions. DNA concentration and purity were assessed using Nanodrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, Germany). DNA was stored at -80 °C until further analyses.

Profiling of DNA methylation levels: Methylation analyses of DNA samples were carried out using Infinium HumanMethylation450 DNA Analysis BeadChip (Illumina), with coverage of more than 480,000 methylation sites. DNA samples were bisulfite-treated using the EZ DNA Methylation kit (Zymo Research Corp, Irvine, CA, USA), and analysed according to the standard protocol for Illumina 450k microarrays, by the Genotyping and Genetic Diagnosis Unit from the University of Valencia, Spain.

Statistics and data analyses

Data extraction: DNA methylation signals from scanned arrays were preprocessed and normalized using the R scripts [21] of Touleimat et al. [50]. After data normalization with the DASEN method [41], a multidimensional scaling (MDS) was carried out to define the source of variation from the pooled β -values of the methylation levels of the samples.

Selection of CpG sites: CpG selection was performed using the “Standard Deviation for β -values” method, which has been shown to be a robust non-specific procedure in the selection of CpG and in the dimension reduction procedure performed in high-throughput databases, prior to cluster analysis [57]. According to this premise, those CpGs with higher standard deviations were chosen, assuming that those CpGs could be expected to have more biological relevance and variability. In addition a concordance correlation analysis [26] was carried out to measure the agreement between DNA methylation levels in blood and buccal swabs. This method is based on a coefficient correction that rewards the linear fit with origin in 0 and a slope with value 1. This analysis is widely used in the validation of different techniques, giving higher coefficients to those techniques presenting best accuracy and precision, with 1 being the value for perfect fit between measurements. The correlation matrix of these selected CpG was plotted in a heatmap, after unsupervised linkage clustering using Euclidean distances, to visualise the methylation patterns of those CpG with better fit between tissues.

Correlation with phenotypical characteristics: Pairwise Pearson and Spearman correlation coefficients between CpGs of both tissues and phenotypical characteristics were estimated to determine the relationship between methylation levels with anthropometrical and biochemical variables. Those CpGs that show association with any phenotypical variable and passed a p-value threshold of 0.01 (which correspond with the minimum common p-value possible for both tissues and tests) in both tissues for Pearson and Spearman tests, were selected for further investigation (Supplementary Figure 1).

The Shapiro-Wilk normality test was used to determine the distribution of phenotypic variables. Accordingly, Student t-test or Mann-Whitney U test were used for parametric and nonparametric continuous variables, respectively. R-Studio software was used to carry out all the analysis and graphics.

Pathway analysis

The Ingenuity Pathway Analysis (IPA) (IPA software, Ingenuity Systems; Qiagen; Redwood City, CA, USA) system was used to analyse functions potentially involved in obesity or non-communicable metabolic disorders. Pathway analysis was conducted on the list of genes from selected CpGs with best fitting in methylation level between tissues, to determine those pathways or functions that could be equally modified in both tissues. Each gene from the list was classified by IPA, according to pathways and functions. Fisher's exact test was used to estimate the likelihood and the p-value of those genes participating in each pathway.

Ethics

The present investigation was approved by the Research Ethics Committee of the University of Navarra (ref. 041/2012). The protocol and all the procedures used during this trial are in agreement with the Helsinki Declaration articles, and European and National guidelines for Good Clinical Practices.

Results

General characteristics: The sample (n=10) was composed of 6 females and 4 males, with a mean age of 36 ± 5 years, and a BMI of 24.2 ± 3.7 kg/m². When the sample was dichotomised into those participants who were normal weight or overweight, defined by the BMI (Table 1), statistically significant differences were found in leptin (6.57 ± 5.93 mg/dl vs. 16.68 ± 8.25 mg/dl, respectively) and triglycerides (47.25 ± 9.07 mg/dl vs. 85.67 ± 25.75 mg/dl, respectively).

Identification of concordant CpG sites: Methylation levels of all CpGs sites were plotted (Figure 1a) to illustrate the concordance in methylation between both tissues. Those CpGs with stronger correlations appear on, or close to, the diagonal line, while those that are more distant from the diagonal region are more discordant. Subsequently, multidimensional scaling (MDS) analysis was carried out to confirm the presence of different methylation patterns for each tissue; the graphical representation for the two main dimensions showed the tissue grouping (Figure 1b).

According to the designated criteria for the CpG selection, the standard deviations for each tissue were plotted to determine the β -value range for the collection (Figure 1c and 1d). Those CpGs with a standard deviation higher than 0.2, in at least one of the tissues, were used for subsequent analyses. In order to prevent discordances between tissues, concordance correlation analysis was carried out. The CpGs with higher concordance coefficient (higher than 0.9) were those which best fitted the diagonal line correlation between tissues. Finally, a total of 277 CpGs met both criteria and were selected for further analysis (Sup. Table 1). The hierarchical cluster analysis of those selected CpGs (Figure 2) exhibited a robust mirroring between tissues in each volunteer, showing similarities in methylation levels (expressed by the colour range) between the paired samples for each volunteer.

Enrichment analysis: Enrichment analysis was undertaken to examine the biological significance of those selected CpGs with higher parallelism between tissues, and to identify pathways or functions associated with obesity and related disorders. After the enrichment analysis with Ingenuity software, the biological pathways statistically significant containing the genes of the selected CpGs, in relationship with metabolic disease linked with obesity, were examined (Table 2). The main metabolic pathways identified by this were those related to glucose metabolism (glucose metabolism disorder, 26 molecules and $p=0.001$; diabetes mellitus, 22 molecules and $p=0.002$; carbohydrate storage, 14 molecules and $p=0.002$). Furthermore, pathways related to obesity (11 molecules and $p=0.007$) and to lipid peroxidation (5 molecules and $p=0.004$) were identified.

Correlation analyses with phenotypic characteristics: Correlation analyses were run between the selected CpGs and the phenotypic data. Some CpGs had high correlation with BMI, HOMA-IR and leptin with a p-value for Pearson and Spearman correlation test less than 0.01 in both tissues (Figure 3). These include: cg01055691 in *GAP43* with BMI ($r=-0.92$ & $\rho=-0.84$ for blood; $r=-0.89$ & $\rho=-0.83$ for buccal sample); cg00095677 in *ATP2A3* with HOMA-IR ($r=0.82$ & $\rho=-0.84$ for blood; $r=-0.8$ & $\rho=-0.83$ for buccal sample); and cg14464133 in *ADARB2* with leptin ($r=-0.9182$ & $\rho=-0.94$ for blood; $r=-0.893$ & $\rho=-0.79$ for buccal sample). A complementary analysis was carried out with all the CpGs included in the three selected genes which presented association with the traits. Only two additional CpGs showed correlation between phenotypical variables when the p-value threshold was reduced to 0.05 for both tests and tissues (cg20473595 in *ATP2A3* for HOMA index and cg16957758 in *ADARB2* for leptin levels)

Discussion

In agreement with previous studies, the current investigation identified specific CpG sites which showed similar methylation patterns in both tissues and which were related to obesity

and metabolic functions, as corroborated with the enrichment software IPA [23]. Methylation of the gene *GAP43* (Growth Associated Protein 43) was inversely correlated with BMI. The protein encoded by *GAP43* is involved in neuronal regeneration and plasticity, but studies in rats have shown that obesity or diabetes can affect its expression [4, 52]. We observed a statistically significant correlation between circulating leptin levels and methylation of a CpG located in double-stranded RNA-specific editase B2 (*ADARB2*) gene which is associated with longevity in humans [40, 46]. Polymorphisms in *ADARB2* are associated with higher abdominal circumference, BMI and triglyceride levels [40], but also with lower adiponectin concentration [39, 40]. Accordingly, Oguro et al proposed that changes in *ADARB2* expression in the thalamus and amygdala could trigger changes in feeding behavior that promote visceral fat accumulation [40]. Finally, a correlation between the HOMA index and the methylation levels of a CpG site located in the gene body of *ATP2A3* was found. This gene encodes for one of the SERCA Ca(2+)-ATPases that has been described to be involved in calcium homeostasis in pancreatic cells [51]. Also, methylation changes in CpGs of this gene have been recently described to be associated with changes in insulin in muscle samples from non-diabetic subjects. [36]. Furthermore, changes in the expression of *ATP2A3* isoforms has been also related with endoplasmic reticulum stress and changes in cell adhesion capacity [11].

Some studies suggest the feasibility of using epigenetic markers as an early biomarker of increased risk of specific diseases [6, 15]. Given the worldwide increase in prevalence of obesity and of obesity-related complex disorders, such as impaired glucose homeostasis, dyslipidaemias, or hypertension, which are modulated by both genetic and environmental factors, further research on early biomarkers may help to monitor and screen for the onset and progress for these diseases [7].

In response to those demands, some studies are already linking epigenetic marks and obesity [2, 12, 16, 34, 43]. However, the routine collection of appropriate samples from specific tissues

for determining the developmental status of these disorders, may be a difficult task in normal clinical practice [5]. For these reasons, there is a need to identify alternative, non-invasive sample sources on which accurately reflect related methylation marks. When such tissues have been identified and validated, this biomarker approach may be applied in screening and monitoring of obesity-related disorders. Consequently, the determination of epigenetic marks that are modified in other non-specific tissues provides the opportunity of delivering personalised procedures in clinical practice [47]. Indeed, some authors have proposed the idea of using these epigenetics marks as a screening tool, to predict the risk of future disease [17, 22]. For the application of these biomarkers in the clinical practice, it would be convenient that they are obtained from non-invasive procedures (such as buccal swabs), if it is confirmed that these mirror the epigenetic markers in the target tissues. In this context, differences in methylation patterns between tissues need to be further investigated [28, 32].

The development of studies that provide evidence to support the use of epigenetic marks in specific tissues as a surrogate of epigenetic status in non-invasive tissues will improve the efficiency of clinical practice [5]. In this aspect, it is also necessary to find tissues with enough cell type homogeneity in order to diminish the effects of confounding changes that could be occurring in tissues such as blood, where the health or diseases status can alter the proportion of white cells present, which may hamper the comprehension of epigenetic marks [31, 49]. For this reasons, our study provides support for the idea that methylation of DNA obtained from buccal cells may be used to identify new markers related to metabolic disorders in a homogenous cell type sample. Findings from the present study also allow identification of alterations in the digestive tract, facilitating investigation of impairments in digestion, absorption and different related disorders, such as inflammatory intestinal disorders [30, 58]. At the same time, some studies have revealed the stability of epigenetic marks over time [17]. Although some studies have reported discordance in DNA methylation results when comparing two accessible human tissues such as peripheral blood mononuclear cells and buccal epithelial

cells [25], other trials have suggested that, for epigenetic population studies, buccal swabs are likely to be a more informative surrogate tissue than blood [27]. In the same way, it is very important to take into account other possible habits that could affect the methylation levels in this surrogated tissues, such as smoking behaviour, which has been shown to be associated with changes in specific methylation sites, in buccal mucosa samples used as a surrogate tissue, that were previously associated with changes in expression of genes related with smoking –related lung diseases [55].

Our findings add to the evidence from other recent studies that aimed to associate epigenetic marks in different tissues with metabolic disorders. For example, Ronn et al [45] tried to determine the effects of age, BMI, and glucose status on DNA methylation and RNA expression levels in adipose tissue and peripheral blood samples. These authors also used a wide coverage array to analyse the methylation levels, and they found a large number of genes whose methylation levels correlated with BMI and HbA1c levels in both tissues. In contrast, other recent epigenome-wide profilings of methylation levels have described different genes that could act as mirror markers in paired blood and adipose tissue samples [24]. However, the authors reported lack of association between DNA methylation levels in blood and weight status, suggesting that blood markers do not mirror adipose tissue markers [24]. Using both blood and buccal samples, Talens et al investigated the relationship between DNA methylation levels and obesity, and proposed that the study of complex disorders could be feasible with well-designed studies [49]. Other researches have used buccal samples in the search for methylation signatures in childhood psychotic symptoms [18], the response to prenatal exposure of maternal cigarette smoking [37] or the symptoms observed in children with fetal alcohol spectrum disorder [42].

The present work proposes the use of surrogate methylation markers in buccal swab samples in relation to obesity, to provide a more accessible tissue than peripheral blood cells to

perform future analyses in clinical practice, with the benefit of no need of venous puncture and the feasibility of storage of the sample before the extraction. The main limitation of the present work is the small number of volunteers, but the wide coverage of the methylation analysis and the selection methods performed are strengths. For these reasons, this study provides the bases for the future development of DNA methylation analyses in buccal swab samples as possible surrogate markers for investigation in obesity and related comorbidities. These findings show the potential application of a non-invasive approach in the identification of surrogate epigenetic biomarkers. Particularly, this study opens the door to further analysis of the roles of *GAP43*, *ATP2A3* and *ADARB2* in relation to weight management.

Acknowledgements

The Food4Me study is supported by the European Commission under the Food, Agriculture, Fisheries and Biotechnology Theme of the 7th Framework Programme for Research and Technological Development, Grant Number 265494. The authors want to thank all the volunteers who took part in the study, as well as Maria Hernández Ruiz de Eguilaz, Salomé Perez-Diez, Blanca Martínez de Morentin and Veronica Ciaurriz for technical and laboratory support. The research leading to these results has received funding from "la Caixa" Banking Foundation (RS-C scholarship). CIBERobn funds are also acknowledged.

1. Agha G, Houseman EA, Kelsey KT, Eaton CB, Buka SL, Loucks EB. Adiposity is associated with DNA methylation profile in adipose tissue. *International journal of epidemiology*. 2015;44(4):1277-87.
2. Aslibekyan S, Demerath EW, Mendelson M, Zhi D, Guan W, Liang L, et al. Epigenome-wide study identifies novel methylation loci associated with body mass index and waist circumference. *Obesity (Silver Spring)*. 2015;23(7):1493-501.
3. Barres R, Kirchner H, Rasmussen M, Yan J, Kantor FR, Krook A, et al. Weight loss after gastric bypass surgery in human obesity remodels promoter methylation. *Cell reports*. 2013;3(4):1020-7.
4. Bekar E, Altunkaynak BZ, Balci K, Aslan G, Ayyildiz M, Kaplan S. Effects of high fat diet induced obesity on peripheral nerve regeneration and levels of GAP 43 and TGF-beta in rats. *Biotechnic & histochemistry : official publication of the Biological Stain Commission*. 2014;89(6):446-56.
5. Burdge GC, Lillycrop KA. Bridging the gap between epigenetics research and nutritional public health interventions. *Genome medicine*. 2010;2(11):80.
6. Burdge GC, Hoile SP, Lillycrop KA. Epigenetics: are there implications for personalised nutrition? *Current opinion in clinical nutrition and metabolic care*. 2012;15(5):442-7.
7. Burdge GC, Lillycrop KA. Environment-physiology, diet quality and energy balance: the influence of early life nutrition on future energy balance. *Physiology & behavior*. 2014;134:119-22.
8. Campion J, Milagro FI, Goyenechea E, Martinez JA. TNF-alpha promoter methylation as a predictive biomarker for weight-loss response. *Obesity (Silver Spring)*. 2009;17(6):1293-7.
9. Campion J, Milagro FI, Martinez JA. Individuality and epigenetics in obesity. *Obesity reviews : an official journal of the International Association for the Study of Obesity*. 2009;10(4):383-92.
10. Celis-Morales C, Livingstone KM, Marsaux CF, Forster H, O'Donovan CB, Woolhead C, et al. Design and baseline characteristics of the Food4Me study: a web-based randomised controlled trial of personalised nutrition in seven European countries. *Genes & nutrition*. 2015;10(1):450.
11. Chaabane C, Corvazier E, Bredoux R, Dally S, Raies A, Villemain A, et al. Sarco/endoplasmic reticulum Ca²⁺ATPase type 3 isoforms (SERCA3b and SERCA3f): distinct roles in cell adhesion and ER stress. *Biochemical and biophysical research communications*. 2006;345(4):1377-85.
12. Choi SW, Claycombe KJ, Martinez JA, Friso S, Schalinske KL. Nutritional epigenomics: a portal to disease prevention. *Adv Nutr*. 2013;4(5):530-2.
13. Collaboration NRF. Trends in adult body-mass index in 200 countries from 1975 to 2014: a pooled analysis of 1698 population-based measurement studies with 19.2 million participants. *Lancet*. 2016;387(10026):1377-96.
14. Cordero P, Campion J, Milagro FI, Goyenechea E, Steemburgo T, Javierre BM, et al. Leptin and TNF-alpha promoter methylation levels measured by MSP could predict the response to a low-calorie diet. *Journal of physiology and biochemistry*. 2011;67(3):463-70.
15. Dick KJ, Nelson CP, Tsaprouni L, Sandling JK, Aissi D, Wahl S, et al. DNA methylation and body-mass index: a genome-wide analysis. *Lancet*. 2014;383(9933):1990-8.
16. Drummond EM, Gibney ER. Epigenetic regulation in obesity. *Current opinion in clinical nutrition and metabolic care*. 2013;16(4):392-7.
17. Feinberg AP, Irizarry RA, Fradin D, Aryee MJ, Murakami P, Aspelund T, et al. Personalized epigenomic signatures that are stable over time and covary with body mass index. *Science translational medicine*. 2010;2(49):49ra67.

18. Fisher HL, Murphy TM, Arseneault L, Caspi A, Moffitt TE, Viana J, et al. Methyloomic analysis of monozygotic twins discordant for childhood psychotic symptoms. *Epigenetics*. 2015;10(11):1014-23.
19. Friedewald WT, Levy RI, Fredrickson DS. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. *Clinical chemistry*. 1972;18(6):499-502.
20. Gemma C, Sookoian S, Dieuzeide G, Garcia SI, Gianotti TF, Gonzalez CD, et al. Methylation of TFAM gene promoter in peripheral white blood cells is associated with insulin resistance in adolescents. *Molecular genetics and metabolism*. 2010;100(1):83-7.
21. Gentleman R, Carey V, Huber W, Irizarry R, Dudoit S. *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*: Springer New York; 2006.
22. Gillberg L, Perfilyev A, Brons C, Thomasen M, Grunnet LG, Volkov P, et al. Adipose tissue transcriptomics and epigenomics in low birthweight men and controls: role of high-fat overfeeding. *Diabetologia*. 2016;59(4):799-812.
23. Guenard F, Tchernof A, Deshaies Y, Perusse L, Biron S, Lescelleur O, et al. Differential methylation in visceral adipose tissue of obese men discordant for metabolic disturbances. *Physiological genomics*. 2014;46(6):216-22.
24. Huang YT, Chu S, Loucks EB, Lin CL, Eaton CB, Buka SL, et al. Epigenome-wide profiling of DNA methylation in paired samples of adipose tissue and blood. *Epigenetics*. 2016;11(3):227-36.
25. Jiang R, Jones MJ, Chen E, Neumann SM, Fraser HB, Miller GE, et al. Discordance of DNA methylation variance between two accessible human tissues. *Scientific reports*. 2015;5:8257.
26. Lin LI. A concordance correlation coefficient to evaluate reproducibility. *Biometrics*. 1989;45(1):255-68.
27. Lowe R, Gemma C, Beyan H, Hawa MI, Bazeos A, Leslie RD, et al. Buccals are likely to be a more informative surrogate tissue than blood for epigenome-wide association studies. *Epigenetics*. 2013;8(4):445-54.
28. Marzi SJ, Meaburn EL, Dempster EL, Lunnon K, Paya-Cano JL, Smith RG, et al. Tissue-specific patterns of allelically-skewed DNA methylation. *Epigenetics*. 2016;11(1):24-35.
29. Matthews DR, Hosker JP, Rudenski AS, Naylor BA, Treacher DF, Turner RC. Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man. *Diabetologia*. 1985;28(7):412-9.
30. McDermott E, Ryan EJ, Tosetto M, Gibson D, Burrage J, Keegan D, et al. DNA Methylation Profiling in Inflammatory Bowel Disease Provides New Insights into Disease Pathogenesis. *Journal of Crohn's & colitis*. 2016;10(1):77-86.
31. McGregor K, Bernatsky S, Colmegna I, Hudson M, Pastinen T, Labbe A, et al. An evaluation of methods correcting for cell-type heterogeneity in DNA methylation studies. *Genome biology*. 2016;17:84.
32. McKay JA, Xie L, Harris S, Wong YK, Ford D, Mathers JC. Blood as a surrogate marker for tissue-specific DNA methylation and changes due to folate depletion in post-partum female mice. *Molecular nutrition & food research*. 2011;55(7):1026-35.
33. Milagro FI, Campion J, Cordero P, Goyenechea E, Gomez-Uriz AM, Abete I, et al. A dual epigenomic approach for the search of obesity biomarkers: DNA methylation in relation to diet-induced weight loss. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*. 2011;25(4):1378-89.
34. Milagro FI, Mansego ML, De Miguel C, Martinez JA. Dietary factors, epigenetic modifications and obesity outcomes: progresses and perspectives. *Mol Aspects Med*. 2013;34(4):782-812.
35. Milagro FI, Martinez JA. Epigenetics of obesity and weight loss. *Endocrinologia y nutricion : organo de la Sociedad Espanola de Endocrinologia y Nutricion*. 2013;60 Suppl 1:12-4.

36. Mudry JM, Lassiter DG, Nylen C, Garcia-Calzon S, Naslund E, Krook A, et al. Insulin and Glucose Alter Death-Associated Protein Kinase 3 (DAPK3) DNA Methylation in Human Skeletal Muscle. *Diabetes*. 2016.
37. Nielsen CH, Larsen A, Nielsen AL. DNA methylation alterations in response to prenatal exposure of maternal cigarette smoking: A persistent epigenetic impact on health from maternal lifestyle? *Archives of toxicology*. 2016;90(2):231-45.
38. Nilsson E, Jansson PA, Perfilyev A, Volkov P, Pedersen M, Svensson MK, et al. Altered DNA methylation and differential expression of genes influencing metabolism and inflammation in adipose tissue from subjects with type 2 diabetes. *Diabetes*. 2014;63(9):2962-76.
39. Nomura K, Eto M, Kojima T, Ogawa S, Iijima K, Nakamura T, et al. Visceral fat accumulation and metabolic risk factor clustering in older adults. *Journal of the American Geriatrics Society*. 2010;58(9):1658-63.
40. Oguro R, Kamide K, Katsuya T, Akasaka H, Sugimoto K, Congrains A, et al. A single nucleotide polymorphism of the adenosine deaminase, RNA-specific gene is associated with the serum triglyceride level, abdominal circumference, and serum adiponectin concentration. *Experimental gerontology*. 2012;47(2):183-7.
41. Pidsley R, CC YW, Volta M, Lunnon K, Mill J, Schalkwyk LC. A data-driven approach to preprocessing Illumina 450K methylation array data. *BMC genomics*. 2013;14:293.
42. Portales-Casamar E, Lussier AA, Jones MJ, MacIsaac JL, Edgar RD, Mah SM, et al. DNA methylation signature of human fetal alcohol spectrum disorder. *Epigenetics & chromatin*. 2016;9:25.
43. Remely M, de la Garza AL, Magnet U, Aumueller E, Haslberger AG. Obesity: epigenetic regulation - recent observations. *Biomolecular concepts*. 2015;6(3):163-75.
44. Ribel-Madsen R, Fraga MF, Jacobsen S, Bork-Jensen J, Lara E, Calvanese V, et al. Genome-wide analysis of DNA methylation differences in muscle and fat from monozygotic twins discordant for type 2 diabetes. *PloS one*. 2012;7(12):e51302.
45. Ronn T, Ling C. Effect of exercise on DNA methylation and metabolism in human adipose tissue and skeletal muscle. *Epigenomics*. 2013;5(6):603-5.
46. Sebastiani P, Montano M, Puca A, Solovieff N, Kojima T, Wang MC, et al. RNA editing genes associated with extreme old age in humans and with lifespan in *C. elegans*. *PloS one*. 2009;4(12):e8210.
47. Stenz L, Zewdie S, Laforge-Escarra T, Prados J, La Harpe R, Dayer A, et al. BDNF promoter I methylation correlates between post-mortem human peripheral and brain tissues. *Neuroscience research*. 2015;91:1-7.
48. Swinburn BA, Sacks G, Hall KD, McPherson K, Finegood DT, Moodie ML, et al. The global obesity pandemic: shaped by global drivers and local environments. *Lancet*. 2011;378(9793):804-14.
49. Talens RP, Boomsma DI, Tobi EW, Kremer D, Jukema JW, Willemsen G, et al. Variation, patterns, and temporal stability of DNA methylation: considerations for epigenetic epidemiology. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*. 2010;24(9):3135-44.
50. Touleimat N, Tost J. Complete pipeline for Infinium((R)) Human Methylation 450K BeadChip data processing using subset quantile normalization for accurate DNA methylation estimation. *Epigenomics*. 2012;4(3):325-41.
51. Varadi A, Molnar E, Ostenson CG, Ashcroft SJ. Isoforms of endoplasmic reticulum Ca(2+)-ATPase are differentially expressed in normal and diabetic islets of Langerhans. *The Biochemical journal*. 1996;319 (Pt 2):521-7.
52. Vo PA, Tomlinson DR. Effects of nerve growth factor on expression of GAP-43 in right atria after sympathectomy in diabetic rats. *Diabetes, obesity & metabolism*. 2001;3(5):350-9.

53. Voisin S, Almen MS, Zheleznyakova GY, Lundberg L, Zarei S, Castillo S, et al. Many obesity-associated SNPs strongly associate with DNA methylation changes at proximal promoters and enhancers. *Genome medicine*. 2015;7:103.
54. Volkmar M, Dedeurwaerder S, Cunha DA, Ndlovu MN, Defrance M, Deplus R, et al. DNA methylation profiling identifies epigenetic dysregulation in pancreatic islets from type 2 diabetic patients. *The EMBO journal*. 2012;31(6):1405-26.
55. Wan ES, Qiu W, Carey VJ, Morrow J, Bacherman H, Foreman MG, et al. Smoking-Associated Site-Specific Differential Methylation in Buccal Mucosa in the COPD Gene Study. *American journal of respiratory cell and molecular biology*. 2015;53(2):246-54.
56. Wang X, Zhu H, Snieder H, Su S, Munn D, Harshfield G, et al. Obesity related methylation changes in DNA of peripheral blood leukocytes. *BMC medicine*. 2010;8:87.
57. Wang X, Laird PW, Hinoue T, Groshen S, Siegmund KD. Non-specific filtering of beta-distributed data. *BMC bioinformatics*. 2014;15:199.
58. Zhang Y, Zhang XR, Park JL, Kim JH, Zhang L, Ma JL, et al. Genome-wide DNA methylation profiles altered by *Helicobacter pylori* in gastric mucosa and blood leukocyte DNA. *Oncotarget*. 2016;7(24):37132-44.

Tables and Figures

Table 1: Anthropometric and clinical characteristics of the study population and comparison by weight status.

	All	Normal weight	Overweight	p
n (female)	10 (6)	4 (3)	6 (3)	
Age (years)	36.1 ± 4.7	35.0 ± 6.1	36.8 ± 4.0	0.453
BMI (kg/m ²)	24.19 ± 3.67	20.72 ± 0.87	26.51 ± 2.76	0.011
Fat mass (%)	26.39 ± 8.70	20.58 ± 9.70	30.27 ± 5.88	0.136
Total Cholesterol (mg/dl)	193.9 ± 47.9	168.7 ± 25.2	210.67 ± 53.9	0.201
HDL-Cholesterol (mg/dl)	58.2 ± 12.2	61.5 ± 4.9	56.0 ± 15.5	0.201
LDL-Cholesterol (mg/dl)	121.6 ± 40.3	97.8 ± 22.1	137.5 ± 43.4	0.201
Triglycerides (mg/dl)	70.3 ± 28.1	47.2 ± 9.1	85.7 ± 25.7	0.011
Glucose (mg/dl)	85.3 ± 4.5	83.8 ± 4.2	86.3 ± 4.8	0.670
Insulin (mU/l)	5.36 ± 3.52	3.84 ± 2.45	6.38 ± 3.95	0.286
HOMA-IR	1.14 ± 0.75	0.81 ± 0.55	1.36 ± 0.83	0.286
ALT (U/l)	25.05 ± 13.37	21.48 ± 7.27	27.43 ± 16.52	0.670
AST (U/l)	20.97 ± 6.40	20.88 ± 7.05	21.03 ± 6.62	0.521
Homocysteine (μmol/l)	10.64 ± 4.40	11.00 ± 3.12	10.40 ± 5.37	0.593
C-reactive protein (μg/ml)	1.45 ± 1.37	0.54 ± 0.60	2.06 ± 1.43	0.136
Cortisol (ng/ml)	164.1 ± 38.1	158.5 ± 31.0	167.9 ± 44.7	0.670
Leptin (ng/ml)	12.64 ± 8.77	6.57 ± 5.93	16.68 ± 8.25	0.033

Values expressed in mean ± standard deviation

Figure 1:

Figure 1. Methylation characteristics by samples: a) Scatter plot of mean methylation levels of >423,000 methylation sites from blood and buccal samples; b) the first two components from MDS analysis carried out from pooled samples; c) histogram of standard deviation distribution of blood samples; and d) histogram of standard deviation distribution of buccal samples.

499 **Figure 2:**

500

501 **Figure 2. Heatmap of selected CpGs:** Hierarchical cluster analysis of methylation levels in pooled samples showing the parallelism
502 in methylation levels between blood and buccal samples of each volunteer in selected CpGs with standard deviation higher than
503 0.2 and concordance coefficient higher than 0.9. Similarities in the colours of paired samples for each volunteer show the
504 analogous methylation levels for the selected CpG in both tissues, confirming visually the results obtained from concordance
505 analysis.

Table 2: Annotated functions of the genes whose methylation levels showed higher parallelism between tissues, according to the enrichment analysis by Ingenuity of selected CpGs with standard deviation higher than 0.2 and concordance coefficient higher than 0.9.

Category	Function Annotation	p-value	Genes	Number of genes
Metabolic Disease	glucose metabolism disorder	0.00108	COL6A1, DIP2C, FASN, GAA, GABRP, GCK, GNAS, HDAC4, HK2, HLA-C, HLA-DQA1, HLA-DQB1, HTR4, HTT, ITGB2, MAGI2, PASK, PDE2A, PTGER3, RNF39, SAMD4A, SCN1A, SLC25A13, TNXB, TRIM26, UBE2E3	26
Endocrine System Disorders	diabetes mellitus	0.00168	COL6A1, DIP2C, FASN, GAA, GABRP, GCK, HK2, HLA-C, HLA-DQA1, HLA-DQB1, HTR4, HTT, ITGB2, MAGI2, PDE2A, PTGER3, RNF39, SAMD4A, SCN1A, TNXB, TRIM26, UBE2E3	22
Carbohydrate Metabolism	quantity of carbohydrate storage	0.0021	FASN, FSTL1, GAA, GCK, GNAS, HDAC4, HK2, HTT, ITGB2, LGALS8, MGAT5B, PCCA, SLC25A13, XYLT1	14
Nutritional Disease	obesity	0.00738	FASN, GABRP, GCK, GNAS, GNRHR, HTR4, PASK, PTGER3, SAMD4A, SCN1A, TXNRD1	11
Carbohydrate Metabolism	metabolism of polysaccharide	0.00221	AOAH, CASR, CEMIP, GAA, GCK, PASK, SULF2, XYLT1	8
Endocrine System Disorders	diabetic complication	0.00599	COL6A1, FASN, MAGI2, PDE2A, PTGER3, SCN1A	6
Carbohydrate Metabolism	quantity of glycogen storage	0.00019	FSTL1, GAA, GCK, GNAS, HK2, PCCA	6
Lipid Metabolism	peroxidation of lipid	0.00374	AATF, HTT, MAPT, PTGER3, SH3BP5	5

Figure 3:

Figure 3: Associations between the methylation levels of CpGs located in *GAP43*, *ATP2A3* and *ADARB2*, in both blood and buccal samples with BMI, HOMA index and leptin levels respectively. The scatter plots represent the methylation level of each CpG which present a p-value <0.01 for both tissues and by Pearson and Searman correlation test. The fitted linear line has been also represented.

Supplementary figure 1:

Supplementary figure 1. Venn diagrams for CpG selection. Venn diagrams depicting the selection of CpGs that correlate in both tissues with significance level $p < 0.01$ in Pearson and Spearman tests with the associated traits. a) Representation of the intersecting number of CpGs that correlates with BMI (kg/m²); the only CpG for this trait was cg01055691 in the *GAP43* gene. b) Representation of the intersecting number of CpGs that correlates with HOMA index; the only CpG for this trait was cg00095677 in the *ATP2A3* gene. Finally, c) Representation of the intersecting number of CpGs that correlates with leptin levels (ng/ml); the only CpG for this trait was cg14464133 in the *ADARB2* gene.

Supplementary table 1: Location characteristics and concordance coefficient between tissues of selected CpGs with standard deviation higher than 0.2 and concordance coefficient higher than 0.9.

ID	Chromosome and position	Gene	Relation to gene region	Relation to UCSC CpG Island	Concordance coefficient
cg00095677	17:3833739	<i>ATP2A3</i>	Body	Open Sea	0.989
cg00156497	6:133663589	<i>EYA4</i>	Body	Open Sea	0.980
cg00211215	6:32552246	<i>HLA-DRB1</i>	Body	Island	0.960
cg00247094	17:77922972	<i>TBC1D16</i>	Body	Island	0.957
cg00256329	17:724374	<i>NXN</i>	Body	Open Sea	0.990
cg00443543	17:1645410	<i>SERPINF2</i>	TSS1500	Open Sea	0.986
cg00540295	9:139616839	<i>FAM69B</i>	Body	N_Shore	0.979
cg00881894	2:166930521	<i>SCN1A</i>	TSS1500	Open Sea	0.950
cg00939438	14:55151579	<i>SAMD4A</i>	Body	Open Sea	0.981
cg01012836	12:125810767	<i>TMEM132B</i>	TSS1500	Open Sea	0.915
cg01055691	3:115376349	<i>GAP43</i>	Body	N_Shore	0.927
cg01081438	1:92417998	<i>BRDT</i>	5UTR	S_Shelf	0.993
cg01128042	10:115465924	<i>CASP7</i>	Body	Open Sea	0.959
cg01296877	1:2274955	<i>MORN1</i>	Body	N_Shore	0.971
cg01502466	6:29898751		Intergenic	S_Shelf	0.975
cg01550813	7:134427910		Intergenic	Open Sea	0.990
cg01584086	11:10373718		Intergenic	Open Sea	0.949
cg01647917	19:542653	<i>GZMM</i>	TSS1500	S_Shore	0.979
cg01808284	5:148031958	<i>HTR4</i>	5UTR	N_Shore	0.969
cg01821635	1:101089253		Intergenic	Open Sea	0.945
cg01835922	11:121626930		Intergenic	Open Sea	0.979
cg01941243	1:227971215		Intergenic	N_Shelf	0.982
cg02093708	16:87577273		Intergenic	Open Sea	0.992
cg02100397	19:646890		Intergenic	Island	0.964
cg02113055	10:45072520		Intergenic	Open Sea	0.985
cg02239586	20:62678536		Intergenic	N_Shore	0.960
cg02299007	8:1140574		Intergenic	S_Shore	0.957
cg02353427	2:25079600	<i>ADCY3</i>	Body	Open Sea	0.985
cg02372404	10:131209556		Intergenic	Open Sea	0.985
cg02464073	21:46349496	<i>ITGB2</i>	TSS1500	N_Shore	0.987
cg02533724	10:130279912		Intergenic	Open Sea	0.927
cg02668233	12:125813494	<i>TMEM132B</i>	Body	N_Shelf	0.993
cg02722613	4:25162898	<i>SEPSECS</i>	TSS1500	S_Shore	0.987
cg02749948	6:32073700	<i>TNXB</i>	5UTR	Open Sea	0.989
cg02890235	19:55735946		Intergenic	Open Sea	0.925
cg02940070	22:43343608	<i>PACSLN2</i>	TSS1500	Open Sea	0.926
cg03028786	2:224897425	<i>SERPINE2</i>	5UTR	Open Sea	0.976
cg03075889	7:1976457	<i>MAD1L1</i>	Body	Island	0.991
cg03084184	10:133981681	<i>JAKMIP3</i>	3UTR	Open Sea	0.946

ID	Chromosome and position	Gene	Relation to gene region	Relation to UCSC CpG Island	Concordance coefficient
cg03115532	6:28185726	<i>LOC222699</i>	Body	N_Shore	0.985
cg03129555	1:198466250		Intergenic	Open Sea	0.994
cg03136151	1:92415048	<i>BRDT</i>	5UTR	S_Shore	0.974
cg03216697	6:31239295	<i>HLA-C</i>	Body	Island	0.974
cg03329597	3:108125523	<i>MYH15</i>	Body	Open Sea	0.937
cg03407524	17:80043349	<i>FASN</i>	Body	S_Shore	0.978
cg03466780	9:140247365	<i>EXD3</i>	Body	N_Shore	0.994
cg03570263	6:30040291	<i>RNF39</i>	Body	S_Shore	0.994
cg03704393	9:97042001	<i>ZNF169</i>	Body	Open Sea	0.995
cg03812172	7:44184403	<i>GCK</i>	3UTR	N_Shore	0.968
cg04003990	2:242056286	<i>PASK</i>	Body	Open Sea	0.946
cg04012354	4:143771325		Intergenic	S_Shelf	0.989
cg04073914	4:84461507	<i>AGPAT9</i>	Body	S_Shelf	0.996
cg04090468	7:156181990		Intergenic	Open Sea	0.965
cg04131969	2:33951647	<i>MYADML</i>	Body	N_Shore	0.993
cg04246708	1:246785611	<i>CNST</i>	Body	S_Shore	0.988
cg04388792	19:12707977	<i>ZNF490</i>	Body	Island	0.992
cg04546413	19:29218101		Intergenic	Island	0.986
cg04875706	15:97311928		Intergenic	Open Sea	0.925
cg04998327	11:103453960		Intergenic	Open Sea	0.985
cg05023192	2:240965916	<i>NDUFA10</i>	TSS1500	S_Shore	0.982
cg05084668	3:125655381	<i>ALG1L</i>	5UTR	Open Sea	0.933
cg05134736	12:378771		Intergenic	Open Sea	0.931
cg05161773	17:75378036	<i>sep-09</i>	5UTR	Open Sea	0.983
cg05291429	17:1494566	<i>SLC43A2</i>	Body	S_Shelf	0.990
cg05308244	12:130968528	<i>RIMBP2</i>	5UTR	Open Sea	0.954
cg05331763	17:80535367	<i>FOXK2</i>	Body	S_Shore	0.988
cg05392448	1:2266933	<i>MORN1</i>	Body	S_Shore	0.928
cg05407200	11:93586725		Intergenic	S_Shelf	0.961
cg05528899	17:57120		Intergenic	Island	0.983
cg05593887	7:77827379	<i>MAGI2</i>	Body	Open Sea	0.961
cg05656210	5:141660565		Intergenic	Open Sea	0.950
cg05708441	2:132186821		Intergenic	S_Shelf	0.989
cg05813498	17:78093353	<i>GAA</i>	3UTR	Open Sea	0.993
cg05867245	20:62402415	<i>ZBTB46</i>	Body	Island	0.984
cg05995465	2:240313602	<i>HDAC4</i>	5UTR	Open Sea	0.983
cg06032337	6:29648468		Intergenic	Open Sea	0.962
cg06051619	10:593275	<i>DIP2C</i>	Body	Open Sea	0.984
cg06378561	3:120151283	<i>FSTL1</i>	Body	Open Sea	0.997
cg06621919	1:208309239	<i>PLXNA2</i>	Body	Open Sea	0.972
cg06684911	19:1792217	<i>ATP8B3</i>	Body	N_Shelf	0.965
cg06894070	11:71238205	<i>KRTAP5-7</i>	TSS200	Open Sea	0.981
cg06899226	15:36524258		Intergenic	Open Sea	0.958

ID	Chromosome and position	Gene	Relation to gene region	Relation to UCSC CpG Island	Concordance coefficient
cg06960717	7:64894780		Intergenic	Open Sea	0.991
cg07456585	14:74704714	<i>VSX2</i>	TSS1500	N_Shore	0.972
cg07844442	10:129144269	<i>DOCK1</i>	Body	Open Sea	0.971
cg07909498	4:79627477		Intergenic	Open Sea	0.950
cg08084984	16:17482912	<i>XYLT1</i>	Body	Open Sea	0.978
cg08146865	3:48335857	<i>NME6</i>	3UTR	Open Sea	0.984
cg08210706	14:95046686	<i>SERPINA5</i>	TSS1500	Open Sea	0.950
cg08238319	5:404713	<i>AHRR</i>	Body	Open Sea	0.962
cg08266644	1:214317404		Intergenic	Open Sea	0.990
cg08292959	17:74878420	<i>MGAT5B</i>	Body	Island	0.989
cg08429705	19:2583601	<i>GNG7</i>	5UTR	S_Shelf	0.977
cg08477332	1:153590243	<i>S100A14</i>	TSS1500	Open Sea	0.976
cg08551408	2:2732624		Intergenic	Open Sea	0.928
cg08717807	16:85531414		Intergenic	S_Shore	0.980
cg08896901	4:152996165		Intergenic	Open Sea	0.947
cg08916385	4:68622534	<i>GNRHR</i>	TSS1500	Open Sea	0.928
cg09015880	17:47207658		Intergenic	N_Shelf	0.981
cg09175792	10:50534598	<i>C10orf71</i>	Body	Island	0.939
cg09281805	7:4751840	<i>FOXK1</i>	Body	Open Sea	0.987
cg09284209	12:124585152		Intergenic	Open Sea	0.924
cg09352518	19:21265421	<i>ZNF714</i>	5UTR	Island	0.949
cg09411910	6:32607224	<i>HLA-DQA1</i>	Body	Open Sea	0.913
cg09438069	4:187581093	<i>FAT1</i>	Body	Open Sea	0.981
cg09518270	1:211308115	<i>KCNH1</i>	TSS1500	S_Shore	0.962
cg09627057	3:15377670	<i>SH3BP5</i>	5UTR	S_Shelf	0.995
cg09866143	3:124861521	<i>SLC12A8</i>	Body	S_Shore	0.989
cg09885502	20:57463991	<i>GNAS</i>	3UTR	Island	0.971
cg10155537	1:248512094	<i>OR14C36</i>	1stExon	Open Sea	0.952
cg10317314	12:31272114		Intergenic	Open Sea	0.944
cg10410513	8:1893887	<i>ARHGEF10</i>	Body	Island	0.986
cg10415021	11:110890228		Intergenic	Open Sea	0.973
cg10526376	2:1788254		Intergenic	Open Sea	0.979
cg10596483	8:143751796	<i>JRK</i>	TSS1500	S_Shore	0.922
cg10610477	2:60712421	<i>BCL11A</i>	Body	Open Sea	0.978
cg10632770	15:81122346	<i>KIAA1199</i>	5UTR	Open Sea	0.981
cg10701801	1:52082593	<i>OSBPL9</i>	TSS200	Open Sea	0.941
cg10760240	17:35414006	<i>AATF</i>	3UTR	Open Sea	0.992
cg10788927	21:47419697	<i>COL6A1</i>	Body	N_Shore	0.970
cg10900271	7:156157852		Intergenic	Open Sea	0.987
cg10909185	11:75139736	<i>KLHL35</i>	Body	Island	0.914
cg10985055	6:30174131	<i>TRIM26</i>	5UTR	Open Sea	0.968
cg10995422	6:32522872	<i>HLA-DRB6</i>	Body	Open Sea	0.979
cg11010744	7:171613		Intergenic	Open Sea	0.972

ID	Chromosome and position	Gene	Relation to gene region	Relation to UCSC CpG Island	Concordance coefficient
cg11418303	1:71511514	<i>PTGER3</i>	Body	N_Shore	0.951
cg11424828	8:2075469	<i>MYOM2</i>	Body	Island	0.913
cg11477175	10:114604922		Intergenic	Open Sea	0.967
cg11663691	4:119759900		Intergenic	S_Shelf	0.982
cg11716267	2:24367828	<i>LOC375190</i>	Body	Open Sea	0.975
cg11738485	19:12877000	<i>HOOK2</i>	Body	Island	0.975
e	10:134800741		Intergenic	N_Shelf	0.903
cg11866463	21:35982790	<i>RCAN1</i>	5UTR	N_Shelf	0.970
cg11888470	6:32526021	<i>HLA-DRB6</i>	Body	Open Sea	0.980
cg12036633	15:63758958		Intergenic	Open Sea	0.961
cg12169700	7:1923695	<i>MAD1L1</i>	Body	Open Sea	0.983
cg12381531	3:8881526		Intergenic	Open Sea	0.959
e	20:62098002	<i>KCNQ2</i>	Body	Island	0.961
cg12421087	10:1083304	<i>C10orf110</i>	Body	Open Sea	0.976
cg12434901	17:61602292	<i>KCNH6</i>	Body	Open Sea	0.978
cg12459000	1:21032072	<i>KIF17</i>	Body	S_Shore	0.921
cg12657416	9:139607421	<i>FAM69B</i>	Body	Island	0.958
cg13127920	12:124810173	<i>NCOR2</i>	Body	N_Shore	0.974
cg13160627	7:211932	<i>FAM20C</i>	Body	Open Sea	0.989
cg13167158	1:1562535	<i>MIB2</i>	Body	N_Shore	0.986
cg13295089	7:155492281	<i>RBM33</i>	Body	Open Sea	0.983
cg13591052	5:97873389		Intergenic	Open Sea	0.965
cg13653328	5:148520669	<i>ABLIM3</i>	TSS1500	N_Shore	0.972
cg13655986	10:129197384	<i>DOCK1</i>	Body	S_Shore	0.974
cg13703663	7:2765455		Intergenic	Open Sea	0.978
cg13748354	7:2289888	<i>NUDT1</i>	Body	N_Shelf	0.985
cg13824456	7:13837775		Intergenic	Open Sea	0.938
e	5:172970	<i>PLEKHG4B</i>	Body	S_Shelf	0.978
cg13972557	5:116075820		Intergenic	Open Sea	0.985
cg14023020	8:1972405		Intergenic	Island	0.942
cg14114910	9:124924045	<i>MORN5</i>	Body	S_Shore	0.981
cg14192979	17:12562530		Intergenic	Open Sea	0.915
cg14195178	17:80849463	<i>TBCD</i>	Body	Open Sea	0.996
e	4:184250312		Intergenic	Open Sea	0.950
cg14252149	1:236707045	<i>LGALS8</i>	Body	Open Sea	0.983
cg14323910	6:32628305	<i>HLA-DQB1</i>	Body	N_Shelf	0.948
cg14350224	11:63885769	<i>FLRT1</i>	3UTR	Island	0.973
cg14464133	10:1245586	<i>ADARB2</i>	Body	Open Sea	0.967
cg14651435	7:157209551	<i>DNAJB6</i>	3UTR	S_Shore	0.977
cg14815891	20:29611903	<i>FRG1B</i>	Body	Island	0.911
cg14883135	2:43051175		Intergenic	N_Shelf	0.965
cg14893161	1:205819251	<i>PM20D1</i>	5UTR	S_Shore	0.944
cg15014361	6:32609212	<i>HLA-DQA1</i>	Body	Open Sea	0.987

ID	Chromosome and position	Gene	Relation to gene region	Relation to UCSC CpG Island	Concordance coefficient
cg15150396	2:71787431	<i>DYSF</i>	Body	Island	0.960
cg15295200	3:139397901	<i>NMNAT3</i>	TSS1500	S_Shore	0.949
cg15613905	5:112454695	<i>MCC</i>	Body	Open Sea	0.972
cg15668967	5:1180364		Intergenic	Open Sea	0.988
cg15775217	12:104676774	<i>TXNRD1</i>	Body	N_Shelf	0.951
cg15800276	7:154958660		Intergenic	Open Sea	0.988
cg15876198	14:95106927	<i>SERPINA13</i>	TSS200	Open Sea	0.976
cg15958422	2:150845309		Intergenic	Open Sea	0.932
cg15988569	10:118380345	<i>PNLIPRP2</i>	TSS200	Open Sea	0.954
cg16069065	7:129412730		Intergenic	Open Sea	0.985
cg16081854	5:308268	<i>AHRR</i>	Body	N_Shore	0.994
cg16112880	1:201123745	<i>TMEM9</i>	TSS200	Island	0.980
cg16144436	17:7645325	<i>DNAH2</i>	Body	S_Shore	0.979
cg16241932	6:157876915	<i>ZDHHC14</i>	Body	Open Sea	0.972
cg16312212	8:18941574		Intergenic	Open Sea	0.985
cg16490124	1:235292369	<i>SNORA14B</i>	TSS1500	Island	0.995
cg16490805	16:10281150		Intergenic	S_Shelf	0.981
cg16527629	6:34524698	<i>SPDEF</i>	TSS1500	Open Sea	0.997
cg16542356	7:1121190	<i>C7orf50</i>	Body	S_Shore	0.967
cg16702660	14:22361282		Intergenic	Open Sea	0.981
cg16792234	7:95828497	<i>SLC25A13</i>	Body	Open Sea	0.940
cg16866567	8:38759261	<i>PLEKHA2</i>	5UTR	Island	0.979
cg16964206	7:91795363	<i>LOC401387</i>	TSS1500	Open Sea	0.972
cg17113602	20:62110277		Intergenic	N_Shore	0.959
cg17149911	2:69812277	<i>AAK1</i>	Body	Open Sea	0.956
cg17181941	2:97004840	<i>NCAPH</i>	Body	S_Shelf	0.955
cg17187785	11:132107476	<i>NTM</i>	Body	Open Sea	0.958
cg17225604	17:22193895		Intergenic	S_Shore	0.955
cg17329602	7:1948947	<i>MAD1L1</i>	Body	N_Shore	0.946
cg17343385	20:62552607	<i>DNAJC5</i>	5UTR	S_Shore	0.988
cg17388779	3:109537321		Intergenic	Open Sea	0.943
cg17671604	19:41032399	<i>SPTBN4</i>	Body	N_Shelf	0.984
cg18032705	20:47253042	<i>PREX1</i>	Body	Island	0.963
cg18075755	4:15429967	<i>C1QTNF7</i>	5UTR	Open Sea	0.958
cg18239511	14:96563269		Intergenic	Open Sea	0.956
cg18346634	1:29190137	<i>OPRD1</i>	3UTR	S_Shore	0.979
cg18827503	13:112574260		Intergenic	N_Shore	0.984
cg18878992	17:43974344	<i>MAPT</i>	5UTR	Island	0.995
cg18938907	1:110230456	<i>GSTM1</i>	1stExon	Island	0.942
cg19235645	12:124787691	<i>FAM101A</i>	5UTR	N_Shore	0.974
cg19248407	7:101662116	<i>CUX1</i>	Body	Open Sea	0.968
cg19311470	4:39460490	<i>RPL9</i>	TSS1500	Island	0.989
cg19353052	12:113516445	<i>DTX1</i>	Body	S_Shore	0.991

ID	Chromosome and position	Gene	Relation to gene region	Relation to UCSC CpG Island	Concordance coefficient
cg19360212	10:123731471	<i>NSMCE4A</i>	Body	N_Shelf	0.972
cg19389973	7:36692197	<i>AOAH</i>	Body	Open Sea	0.990
cg19504860	19:21512191	<i>ZNF708</i>	1stExon	Open Sea	0.986
cg19949776	15:51236304	<i>LOC100132724</i>	TSS200	Open Sea	0.961
cg19996396	3:121946297	<i>CASR</i>	5UTR	N_Shelf	0.976
cg20040891	12:133007812		Intergenic	N_Shore	0.939
cg20090066	11:72327160	<i>PDE2A</i>	Body	Open Sea	0.968
cg20205188	10:1251771	<i>ADARB2</i>	Body	N_Shore	0.987
cg20218571	18:77678514	<i>PQLC1</i>	Body	Island	0.993
cg20536971	13:100989375	<i>PCCA</i>	Body	Open Sea	0.944
cg20592836	20:33292126	<i>TP53INP2</i>	TSS200	Island	0.993
cg20961245	6:29649092		Intergenic	Open Sea	0.920
cg21130926	20:46415320	<i>SULF2</i>	TSS1500	Island	0.989
cg21139150	21:46976175		Intergenic	Island	0.946
cg21209485	1:2529359	<i>MMEL1</i>	Body	S_Shore	0.978
cg21294301	1:8120055		Intergenic	Open Sea	0.984
cg21446655	2:181866505	<i>UBE2E3</i>	Body	Open Sea	0.975
cg21792493	14:104652867		Intergenic	Open Sea	0.918
cg21810411	6:32549139	<i>HLA-DRB1</i>	Body	N_Shelf	0.982
cg21885112	7:150026567	<i>C7orf29</i>	TSS1500	Open Sea	0.965
cg22274196	13:95958190		Intergenic	S_Shelf	0.986
cg22274273	6:135137590		Intergenic	Open Sea	0.950
cg22276800	2:75059013	<i>HK2</i>	TSS1500	N_Shore	0.988
cg22336867	19:39265241	<i>LGALS7</i>	TSS1500	S_Shelf	0.980
cg22901347	3:171138553	<i>TNIK</i>	Body	Open Sea	0.952
cg22984282	6:32632568	<i>HLA-DQB1</i>	Body	Island	0.938
cg23187802	10:81173663	<i>ZCCHC24</i>	Body	Open Sea	0.951
cg23248424	5:179741104	<i>GFPT2</i>	Body	Island	0.930
cg23468456	3:134033339		Intergenic	S_Shore	0.970
cg23489384	14:21199099		Intergenic	Open Sea	0.988
cg23499373	19:58578578	<i>ZNF135</i>	Body	Open Sea	0.960
cg23564471	1:204951209	<i>NFASC</i>	Body	Island	0.952
cg23691406	14:71112909	<i>TTC9</i>	Body	S_Shelf	0.994
cg23758822	17:41437982		Intergenic	N_Shore	0.990
cg23762722	17:43817897		Intergenic	Open Sea	0.964
cg23803868	19:2340235	<i>SPPL2B</i>	Body	Island	0.993
cg23936477	18:73139742	<i>C18orf62</i>	TSS200	Open Sea	0.985
cg24018148	6:39290715	<i>KCNK16</i>	TSS1500	Open Sea	0.970
cg24506221	1:110230401	<i>GSTM1</i>	TSS200	Island	0.968
cg24634471	8:143751801	<i>JRK</i>	TSS1500	S_Shore	0.972
cg24638099	6:32526027	<i>HLA-DRB6</i>	Body	Open Sea	0.969
cg24643105	11:113928473		Intergenic	N_Shore	0.989
cg24859648	20:17680544	<i>BANF2</i>	TSS200	Open Sea	0.950

ID	Chromosome and position	Gene	Relation to gene region	Relation to UCSC CpG Island	Concordance coefficient
cg25013753	10:49654342	<i>ARHGAP22</i>	3UTR	N_Shelf	0.969
cg25174111	11:65631880	<i>MUS81</i>	Body	S_Shelf	0.994
cg25203856	11:7059980	<i>NLRP14</i>	Body	Open Sea	0.966
cg25593194	1:17394374	<i>PADI2</i>	3UTR	Open Sea	0.961
cg25674027	12:103325781		Intergenic	Open Sea	0.986
cg25873514	17:81014413		Intergenic	N_Shore	0.984
cg26035071	1:209982407		Intergenic	S_Shelf	0.979
cg26128129	4:3235589	<i>HTT</i>	Body	Open Sea	0.991
cg26179400	2:153355619	<i>FMNL2</i>	Body	Open Sea	0.954
cg26184765	4:103172826	<i>SLC39A8</i>	3UTR	Open Sea	0.974
cg26422465	1:3561008	<i>WDR8</i>	Body	Island	0.995
cg26813483	13:111980537	<i>C13orf16</i>	Body	S_Shore	0.932
cg26846076	17:20128997	<i>CYTSB</i>	Body	Open Sea	0.947
cg26864661	12:76661181		Intergenic	Open Sea	0.923
cg27056740	14:101507727	<i>MIR300</i>	Body	Open Sea	0.992
cg27086157	13:114872606	<i>RASA3</i>	Body	N_Shelf	0.977
cg27129755	5:170210586	<i>GABRP</i>	TSS200	Open Sea	0.937
cg27225663	6:169142715		Intergenic	Open Sea	0.983
cg27288829	19:3771880	<i>RAX2</i>	5UTR	S_Shore	0.968
cg27433479	1:1687481	<i>NADK</i>	Body	S_Shore	0.991
cg27468880	7:965995	<i>ADAP1</i>	Body	Island	0.969
cg27481428	7:155623690		Intergenic	Island	0.918
cg27586797	5:13664584		Intergenic	Open Sea	0.961

Figure legends:

Figure 1. Methylation characteristics by samples: a) Scatter plot of mean methylation levels of >423,000 methylation sites from blood and buccal samples; b) the first two components from MDS analysis carried out from pooled samples; c) histogram of standard deviation distribution of blood samples; and d) histogram of standard deviation distribution of buccal samples.

Figure 2. Heatmap of selected CpGs. Hierarchical cluster analysis of methylation levels in pooled samples showing the parallelism in methylation levels between blood and buccal samples of each volunteer in selected CpGs with standard deviation higher than 0.2 and concordance coefficient higher than 0.9. Similarities in the colours of paired samples for each volunteer show the analogous methylation levels for the selected CpG in both tissues, confirming visually the results obtained from concordance analysis.

Figure 3: Associations between the methylation levels of CpGs located in *GAP43*, *ATP2A3* and *ADARB2*, in both blood and buccal samples, with BMI, HOMA index and leptin levels respectively. The scatter plots represent the methylation level of each CpG which present a p-value <0.01 for both tissues and by Pearson and Spearman correlation tests. The fitted linear line has been also represented.

Supplementary figure 1. Venn diagrams for CpG selection. Venn diagrams depicting the selection of CpGs that correlate in both tissues with significance level $p < 0.01$ in Pearson and Spearman tests with the associated traits. a) Representation of the intersecting number of CpGs that correlates with BMI (kg/m^2); the only CpG for this trait was cg01055691 in the *GAP43* gene. b) Representation of the intersecting number of CpGs that correlates with HOMA index; the only CpG for this trait was cg00095677 in the *ATP2A3* gene. Finally, c) Representation of the intersecting number of CpGs that correlates with leptin levels (ng/ml); the only CpG for this trait was cg14464133 in the *ADARB2* gene.