ECRONICON OPEN ACCESS

EC NEUROLOGY
Research Article

Blood Samples as a Surrogate for Brain Samples in Methylation Studies

Bediaga NG[1], Elcoroaristizabal X[1], Calvo B[2], Inza I[2], Pérez A[2], Acha-Sagredo A[3], Álvare-Álvarez M[4], Gomez Busto F[5], Artaza Artabe I[6], Sevillano R[7], Lozano JA[2] and Marian M de Pancorb[1]*

[1]BIOMICs Research Group, Lascaray Research Center, University of the Basque Country UPV/EHU, Vitoria-Gasteiz, Spain

[2]Intelligent Systems Group, Donostia - San Sebastián, Basque Country, Spain

[3]Oral Medicine and Pathology, Department of Stomatology II, UFI 11/25, University of the Basque Country (UPV/EHU), Leioa, Spain

[4]Servicio de Genómica: Banco de ADN, University of the Basque Country UPV/EHU, Vitoria-Gasteiz, Spain

[5]Centro Integral de Atención a Mayores (C.I.A.M) "San Prudencio", Integral Senior Care Center (CIAM) "San Prudencio", Vitoria-Gasteiz, Spain

[6]Residencia y Unidad Sociosanitaria Orue, Amorebieta, Vizcaya, Spain

[7]Basque Institute of Legal Medicine, Donostia - San Sebastián, Basque Country, Spain

***Corresponding Author**: Marian M de Pancorbo, BIOMICs Research Group, Lascaray Research Center, University of the Basque Country UPV/EHU, Vitoria-Gasteiz, Spain.

Received: February 17, 2017; **Published:** March 20, 2017

Abstract

Background: Both brain and blood can respond to certain environmental stimuli through epigenetically-mediated changes and several lines of evidence suggest that some of these changes are indeed to some extent concordant between both tissue types. Identifying epigenetic biomarkers for which measurement in the blood can be representative of their methylation levels in the brain can be of paramount importance, especially in studies of neurologic or psychiatric traits, where it is unfeasible to directly access the target organ-the brain-in live individuals. However, only a very limited number of investigations have compared the methylation DNA patterns of blood and brain tissue taken from the same healthy human individuals, and this is determinant to measure the intra-individual DNA methylation concordance across different tissues.

Results: We analyzed the DNA methylation profile of 27,579 CpG dinucleotides of two different brain regions (cortex and hippocampus) and blood from 10 autopsy subjects in order to obtain an inventory of epigenetic marks for which measurements in blood are informative for the brain tissue. In addition, we have developed an R application, "Epibrain", which assists researchers in finding the best subset of blood-based surrogate biomarkers to predict methylation status in the brain.

Conclusion: Our results suggest that the error we make when taking a blood sample instead of a brain sample is, in general, greater that the variance we observe between individuals. However, there is a small proportion of CpG for which methylated measurements in blood are representative of methylation in brain, indicating that blood could be used as a surrogate for brain tissue at a small number of CpG sites whose methylation changes in blood could be a good indicator of epigenetics changes induced by pathological alterations in brain.

Keywords: *DNA Methylation; Brain; Blood; Surrogate Tissue*

Abbreviations

GEO: Gene Expression Omnibus; PCA: Principal Components Analysis, TSS: Transcription Start Site; CGI: CpG Islands; AD: Alzheimer Disease

Background

DNA methylation is a key epigenetic mechanism that plays a crucial role in transcriptional regulation and controls several neurobiological and cognitive processes from neurogenesis and brain development [1] to neuronal activity [2], memory formation [3], drug addic-

tion [4] and neurodegeneration [5]. In fact, altered DNA methylation profiles have been associated with a variety of human neurological disorders such as Alzheimer's and Parkinson's diseases [6], Rett syndrome [7] or schizophrenia and bipolar disorders [8,9].

DNA methylation patterns are variable between cells and tissue types and fluctuate in time according to a number of factors such as age-related [10], environmental [11] or nutritional factors [12] among others. The dynamism and tissue-specificity of epigenetic marks make the study design a challenge when relevant tissue is not accessible from living individuals, being this particularly true for studies of neurologic or psychiatric traits. Thus, it is crucial for current epigenetic studies to address the extent to which easily accessible peripheral tissues can be used to ask questions about the epigenomic changes taking place in inaccessible tissues such as the brain. In this regards, several lines of evidence suggest that both brain and blood can respond to certain environmental stimuli trough epigenetically-mediated changes and that these changes are indeed to some extent concordant between both tissue types [13-19]. However, many of these studies are based on animal models rather than on human beings and those that have been undertaken in human beings usually interrogate tissues from multiple individuals and thus, it is often impossible to determine intra-individual DNA methylation variation across different tissues. To our knowledge, there are just a few studies [13:15] where an extensive characterization of the methylation patterns of brain and blood taken from the same individuals has been performed. Hannon., *et al.* [14] measured DNA methylation in matched DNA samples isolated from whole blood and different brain regions from a cohort of 122 individuals with Alzheimer's disease neuropathology and healthy controls, in order to explore the co-variation between tissues as well as the extent to which methylomic variation in blood is predictive of interindividual variation identified in the brain. Walton., *et al.* [15] also assessed the correspondence of DNA methylation between matched living brain tissue and blood samples obtained during neurosurgical treatment of 12 patients with temporal lobe epilepsy. These are two very valuable studies that help us elucidate the utility of blood for the study of disorders that primarily manifest in the brain, however, most of the individuals studied were not healthy individuals but rather individuals with neurological disease, and thus there is a clear lack of information about the brain-blood epigenomic correspondence in healthy individuals.

In the current study, we analyzed the DNA methylation profile of 27,579 CpG dinucleotides of two different brain regions (cortex and hippocampus) and blood from 10 autopsy subjects that were free of any neuropathology disease and created an inventory of CpG for which measurement in the blood can be representative of their methylation status in the brain. In addition, we have created an R-based application where users can filter the CpGs based on their methylation status in brain and blood tissue, concordance between the two brain regions and blood, interindividual variation or presence of polymorphic CpGs, SNPs at the probe or cross-reactiveness.

Methods

Ethics statement

This study counts with a favorable ethical inform from the Faculty of Pharmacy of the University of the Basque Country, signed on 26th September 2008. Written informed consent was obtained from the next-of-kin to postmortem individuals where it was stated that while the deceased was alive he/she had never expressed an objection to having a post mortem or tissue being used for non-diagnostic purposes when they died. The research was carried out according to the Declaration of Helsinki.

Sample collection and DNA preparation

Brain tissues and whole blood samples were obtained from the Basque Institute of Legal Medicine (IMLV, Donostia-San Sebastian, Spain). For this study, 10 post-mortem brain samples (all free of neuropathological and neuropsychiatric disease) and their paired blood samples were collected at the time of autopsy and stored at -80ºC. Different brain regions were dissected from each sample by a trained neuropathologist. All the samples used in the methylomic profiling but one belonged to healthy individuals who died of some traumatic event. In addition, an independent set of matched frontal cortex and cerebellum samples was obtained from 33 additional individuals for the technical validation of the informatic tool developed herein. A detailed list of the samples used for methylomic profiling in this study is given in Table S1.

Sample ID	Tissue type	Age at time of death	Sex	Death cause	Any associated disease	Post-mortem Delay (hrs)
2	C;H;B	73	Male	Traumatic	No	13.2
4	C;H;B	73	Male	Traumatic	No	22.7
7	C;H;B	84	Male	Traumatic	No	14.5
9	C;H;B	69	Female	Traumatic	No	20.0
47	C;H;B	85	Male	Suffocation	No	19.9
57	C;H;B	87	Female	Traumatic	No	7.8
60	C;H;B	76	Male	Suffocation	No	14.0
61	C;H;B	80	Female	Cardiorespiratory failure	Heart disease	17.2
64	C;H;B	85	Male	Traumatic	No	21.5
71	C;H;B	79	Male	Suffocation	No	10.5

C: Cortex; H: Hippocampus; B: Blood

Table S1: *List of samples used for the DNA methylome profiling.*

DNA was extracted from frozen tissue specimens using the QIamp DNA investigator Kit (Qiagen). The DNA was quantified spectrophotometrically (NanoDrop® ND1000; Thermo Fisher Scientific Inc., Waltham, MA, USA) and purity was determined by ratios of absorvanceee260/280 and 260/230. Bisulfite conversion of the genomic DNA samples (1 µg each) was carried out using the EZ DNA Methylation™ Gold kit (Zymo Research, Orange, CA, USA) according to the manufacturer's protocol.

Methylation analysis with the Illumina Infinium HumanMethylation27K BeadChip

DNA methylation analysis of a total of 30 tissue samples was performed with the Illumina HumanMethylation27 DNA Analysis BeadChip assay, (Illumina, San Diego, CA) according to the manufacturer's standard protocols. This BeadChip interrogates 27,578 CpG sites from 14,495 protein-coding gene promoters and 110 microRNA gene promoters at a single-nucleotide resolution.

Computational analysis

Data preprocessing and quality control

Initial array image processing and quality control was performed using the GenomeStudio Methylation Module. The methylated and unmethylated measured intensities were loaded and preprocessed using the lumi package available at Bioconductor repositories (Du., *et al*. 2008). Data preprocessing included background adjustment, color channel adjustment, variance stabilization and quantile normalization. The beta-values were estimated as M/(M+U+100), with M being the (normalized) intensity of the methylated probes and U the corresponding intensity of the unmethylated ones. As an additional quality control of the data, we have constructed an artificial reference array with the median value in the public datasets for each site and compared the beta values measured in our samples with the median value obtained from public datasets (135 whole blood, 112 brains and 99 PBMCs). By doing so, we concluded that none of our arrays has to be labeled as an outlier.

Once data quality was controlled and preprocessed, methylation data was treated as both discrete and continuous data. For the discretization, we divided the methylation levels into three groups: demethylated (0 - 0.25), intermediate (0.25 - 0.75) and methylated (0.75 - 1).

Estimation of the intra-individual variation (Error)

We use the term error to refer to the error we would make if we assume that methylation level/status in blood is the same as in brain. Therefore, the higher the intra-individual cross-tissue (i.e. brain-blood) variation, the higher the error. We measured the error for both

continuous and discrete versions of the data. Error based on discretized data was based on the ratio of paired samples for which the state in the cortex was different from the state in blood given any CpG site. The continuous version of the error was measured as the ratio between the absolute difference between the beta-value in both tissues (cortex and blood, for instance) and the mean beta-value for each site and each individual. Next, the error associated with each site was computed as the median error obtained for all of the individuals.

Estimation of the inter-individual variation

Inter-individual variation was measured in terms of the standard deviation when considering the methylation levels (continuous) and by means of Shannon's entropy when considering the methylation status (dichotomous). The latter was computed using the entropy function available in R's info they package [30]. We considered that 10 samples are not enough to predict the inter-individual methylation variation, and thus, inter-individual variation was measured in two different fashions. The first one reflected the variation of any one CpG across the interrogated 10 samples for anyone tissue type (i.e. "SD cortex" or "Entropy cortex", for the continuous and discrete data respectively). This variation is not commented in the text but it is available in the bioinformatics tool developed herein. The second considered publicly available DNA methylation data from cortex tissue (i.e. "SD cortex all" or "Entropy cortex all", for the continuous and discrete data respectively). Inter-individual variation in the text only refers to this second fashion. The data used were downloaded from the GEO (accession number GSE36194). This Illumina DNA methylation array dataset (n = 724), assess DNA methylation levels in frozen brain tissue from the cerebellum and matched frontal cortex tissue from 318 subjects. Only data from the cortex tissue was employed for our analysis (n = 370).

Estimation of the cross-tissue variability

Cross-tissue variability at both continuous ("SD Mixed Tissues") and discretized ("Entropy Mixed Tissues") data levels was measured in publicly available data of tissue-derived from each of the three embryonic layers. Data were downloaded from the GEO database and corresponded to the following series: (i) GSE54025 and GSE30090 contain methylation data from colon, lung and thyroid tissue, all derived from the endoderm, (ii) GSE49909 corresponds to kidney, buffy coat and muscle, derived from de mesoderm and (iii) GSE49905, GSE28368 and GSE31835 contain methylation data from brain, nasal epithelium and skin originated from the ectoderm. Each of the three groups corresponding to the three embryonic layers were mixed, and variability was measured as described in the above section (inter-individual variability in cortex).

Summarization of the global methylation levels

Given the (usual) bimodality of the distribution of the beta-value, we estimated the probability distribution of the beta-value (i.e., their density) by means of kernel-based density estimators. We computed the probability of having a demethylated site and a methylated site. These probability values were calculated using the density setting with thresholds at 0.25 and 0.75. Thus, we computed, from each density, the probability of having a site with a beta-value lower than 0.25 and the probability of having a site with a beta-value above 0.75.

Definition of the cross-reactive or polymorphic probes

Based on lists of cross-reactive and polymorphic probes that Chen., *et al.* [31] published for the Infinium HumanMethylation450K, we have adapted this list and defined which of the 27,578 probes included in the InfiniumHumanMethylation27K were cross-reactive or polymorphic and therefore could bias the methylation analysis. Over 94% of loci (25,978) present on the HumanMethylation27 array were included in the HumanMethylation450 array content as previously stated in bibliography [29]. Among these 25,978 loci in the Human Methylation27 array, 2.4% were cross-reactive, 7.4% were polymorphic and 38.8% of the probes overlapped at least one SNP **(Table S2).**

Data access

The data used in this study have been packaged and deposited in NCBI's Gene Expression Omnibus (GEO) repository and are accessible through GEO Series, accession number GSE57300 (Reviewer link: http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=kbmjesisrlohped&acc=GSE57300).

Interactive Web application: Epibrain

EpiBrain is implemented as a Shiny application within R. It contains the DNA methylation data from 10 matched blood and brain (cortex and hippocampus) samples plus additional 7 publicly available Illumina DNA methylation array data sets (n = 676) involving 9 healthy tissue datasets. Based on these methylation data, the program calculates a number variables for any given CpG in the array such as the error made using blood as a surrogate of brain in DNA methylation studies, methylation level and status in blood and brain, inter-individual variation across blood or brain tissues as well as other statistical estimations. Additionally, the program also provides information about the location, distance to TSS, non-specific probes (cross-reactive), polymorphic CpGs or presence of at least one SNP at the probe based on the list that Chen et al published for the Infinium HumanMethylation450K [31]. Methylation data are treated both as continuous and categorical values, and all tables generated by EpiBrain can be downloaded as csv files by clicking on the "Save Data as CSV File" button in the upper part of the application.

How to run EpiBrain

1 Installation of R and RStudio
An installation guide for R and RStudio is provided by Torfs and Brauer [33].
2 Start the tool – graphical interface
In case of the first time Epibrain is run, open RStudio and run following commands:

```
install.packages("devtools")
library (devtools)
devtools::install_github ("shiny", "rstudio")
runApp("EpiBrain/")
```

In next applications, the commands can be limited to following lines:

```
library (devtools)
devtools::install_github ("shiny", "rstudio")
runApp("EpiBrain/")
```

Results

Epibrain: an interactive bioinformatic tool for determining the best surrogate CpG sites in blood

Using our matched brain-blood DNA methylation datasets, as well as 7 additional publicly available Illumina DNA methylation array data sets (n = 676) involving 9 healthy tissue datasets, we developed a software package called Epibrain (http://www.sc.ehu.es/ccwbayes/members/borxa/EpiBrain.zip). Epibrain uses R, an open source statistical language and provides a graphical user interface, with a menu-driven format, that facilitates researchers working with the results of our comparisons. The main goal of the software is to show the error made using blood as a surrogate of the brain in DNA methylation studies for any CpG loci included in the HumanMethylation27K microarray. Additionally, the software shows the methylation status and inter-individual variability of each CpG in both blood and brain tissue and contains some extra information such as the location, distance to TSS, non-specific probes (cross-reactive), polymorphic CpGs or presence of at least one SNP at the probe.

Users may filter the CpGs by defining maximum and minimum values for any of the variables mentioned above (error, methylation status, cross-reactivity…). Of note that in order to have a more accurate estimation of the variability, interindividual variation is based on both our cohort of 10 cortex samples and a publicly available set of 370 cortex samples (GSE36194). Inter-individual variation was computed for each age range (young, middle and young) although we found that there were not major inter-individual differences among age ranges.

DNA methylation patterns are tissue specific but are largely conserved across tissues

Principal Components Analysis (PCA) of the methylation data showed that there are remarkable tissue-specific differences in DNA methylation, with a clear difference between the brain tissues (cortex and hippocampus) and blood (Figure 1). However, comparison of DNA methylation patterns across tissue types revealed that many of the CpG sites (39.5% and 68.9% of the CpG sites based on continuous and dichotomized data respectively) displayed no major methylation differences across the tissues within an individual (i.e the blood-brain intra-individual variation or error < 0.01), thus indicating that DNA methylation values are largely conserved across these tissues.

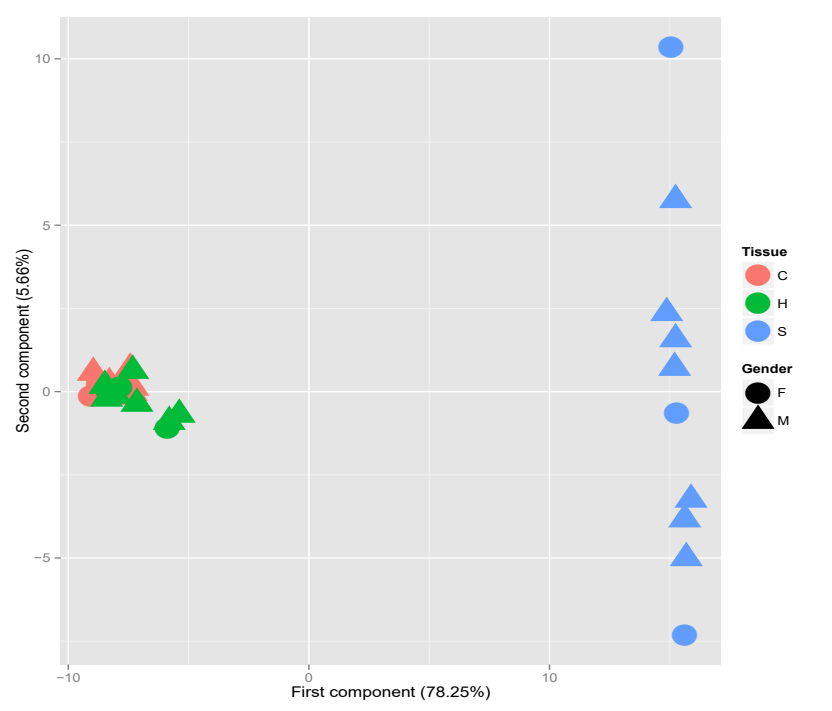


Figure 1: *PCA-based clustering of the samples. Each individual sample was plotted according to the first two components (that account for roughly 80% of the variability). As seen in this figure, the first component perfectly separates brain samples (cortex (C) and hippocampus (H) in the left side of the plot) and blood samples (surrogate whole blood (S) in the right side). The second component further clusters brain and blood samples, but there is not a perfect separation of either the cortex and hippocampus.*

For the majority of DNA methylation sites, interindividual variation in brain greatly exceeds intraindividual cross-tissue (blood-brain) variation

For most of DNA methylation sites, intra-individual brain-blood variation (measured in error) increases as the inter-individual cortex variation augments (Figure S1), and thus methylomic variation in blood is not predictive of interindividual variation in the brain. By us-

ing EpiBrain, we have identified CpGs showing low intra-individual blood-brain variation rates (error < 0.01) and filter out those CpGs displaying similar brain methylation levels between individuals (SD > 0.1) as well as those at the X chromosome (since inter-individual variations at this chromosome might reflect gender variation rather than true inter-individual variation). Only 39 DNA methylation sites out of the whole set of 27,579 fulfilled this criterion. Interestingly, some of these top ranked CpGs were located at genes that have already been associated with neurodegenerative diseases such as THRSP [21,22], GMR4 [23] or B3GALT1 [24,25], and there was not a significant overrepresentation of any of the gene regions (Figure S2).

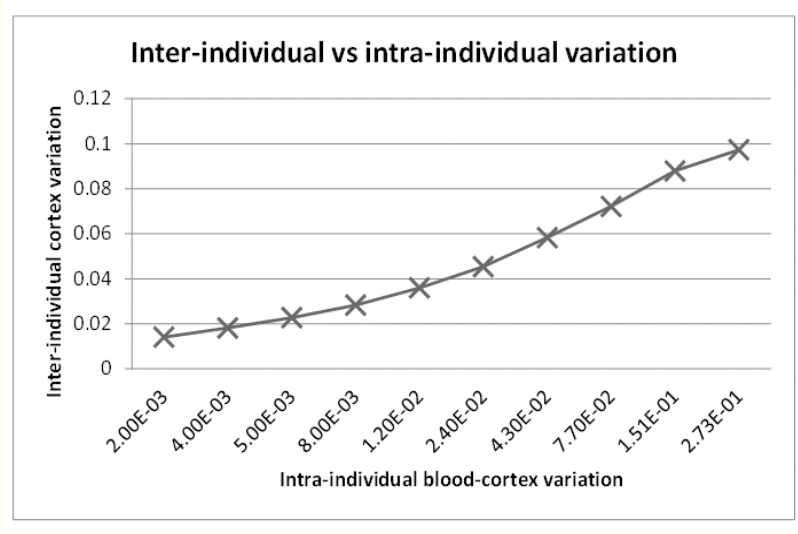


Figure S1: *Shows the inter-individual cortex methylation as a function of the error. It can be seen how intra-individual blood-brain variation (measured in error) increases as inter-individual cortex variation becomes greater.*

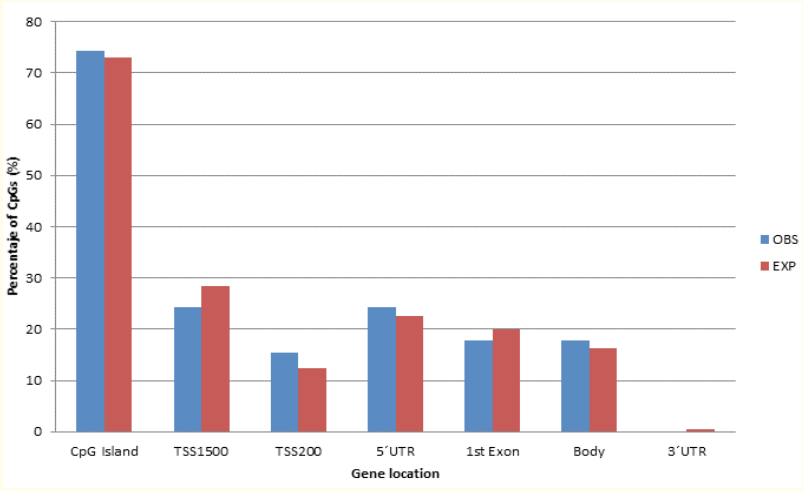


Figure S2: *Representation of the gene locations for the top-ranked probes (i.e. low intra-individual blood-cortex variation and high inter-individual variation) (OBS) in comparison with the gene locations of the CpGs in the Illumina HumanMethylation27 (EXP).*

Postmortem delay does not have a significant impact on the DNA methylation profile

In order to assess the impact that the postmortem delay could have on the DNA methylation profile, we calculated the correlation between the postmortem delay and methylation levels at any given CpG sites for each tissue type, by calculating Spearman's correlation coefficient and subsequent significance testing. After adjusting p-values for multiple comparisons (Benjamini and Hochberg), we found that there were just two CpG sites showing a significant association with the postmortem delay cg03764506 (p-value < 0.05) and cg13433272 (p-value < 0.05), thus concluding that postmortem delay did not have a significant impact on the DNA methylation profile within the range of time in our data.

Brain to blood surrogate epigenetic marks are also informative for other tissues

It is well established that certain CpGs share the same DNA methylation profiles across tissue types [20]. In addition, it has been reported that DNA methylation changes occurring in response to environmental stimuli are not stochastically distributed across the genome and that these changes are indeed to some extent concordant between in rhesus macaque prefrontal cortex and T cells [17]. In order to study the extent to which those marks that show a high epigenetic brain to blood equivalence might be also informative for other internal tissues, we have computed a new variable named "SD Mixed Tissues", which reflects the DNA methylation variability across more than nine different tissues coming from the 3 germ layers. As expected, we found a significant association between the intra-individual variation between blood and brain and "SD Mixed Tissues" (Pearson = 0.7, p-value = 2.2e-16), with more than 95% of CpGs showing low intra-individual cortex-blood variation (error < 0.01, blood-cortex) displaying, in turn, a low cross-tissue variation (SD Mixed Tissues < 0.1). Therefore, we can hypothesize that the methylation of those CpGs showing a high blood to brain equivalence will also be informative for other internal tissues.

DNA methylation status and brain to blood correlation differs across different gene locations or CpG islands

With respect to the DNA methylation status across different gene locations or CpG islands (gene promoter, Transcription Start Site (TSS), 3' UTR, first exon, body of the gene and 5' UTR), we observed that DNA methylation profile does not differ across samples, with a high density of low methylated sites (< 0.25) around the start site (TSS200, 1st exon, and 5´UTR) and an elevated proportion of medium and high methylated CpG sites in the body and 3´UTR of the gene (Figure 2).

On the other hand, we also studied both the intra-individual and inter-individual variations across the different annotated genetic regions sites and observed significant differences among the genetic regions. The hippocampus- and cortex-blood intra-individual variation (measured in error) was significantly lower within CpG Islands (CGI) than outside CGIs (0.04 vs 0.16 error rate, $p < 0.001$). Regarding the error rate relative to the distance to the TSS, we observed that the intra-individual cross-tissue (blood-brain) variation decreases near the TSS, and that it increases as the CpG site moves away from the TSS (Figure 3A), being the intra-individual blood-brain variation higher at those CpGs in the 3´UTR (0.11 error rate), followed by those at the TSS1500 and body of the gene (0.074 and 0.076 of error rate respectively), and finally those in the TSS200, 5´UTR and first exons (0.065, 0.066 and 0.067 respectively) (Figure 3B). However, it is worth pointing out that inter-individual and intra-individual variations seem to follow similar patterns of variation across the different genetic regions (Figure 4). For instance, CpGs in 3´UTR, which generally show high inter-individual brain variations (i.e. the ones that are in principle more interesting from the epidemiologic point of view), tend to display high intra-individual blood-brain differences, and thus blood could not be used to predict its methylation in the brain. On the other hand, CpGs with low intra-individual between blood and brain variations, such as the ones in the TSS200, 5´UTR and first exons, exhibit in turn very low inter-individual cortex variation, and thus they are not so interesting from the epidemiologic epigenetic point of view.

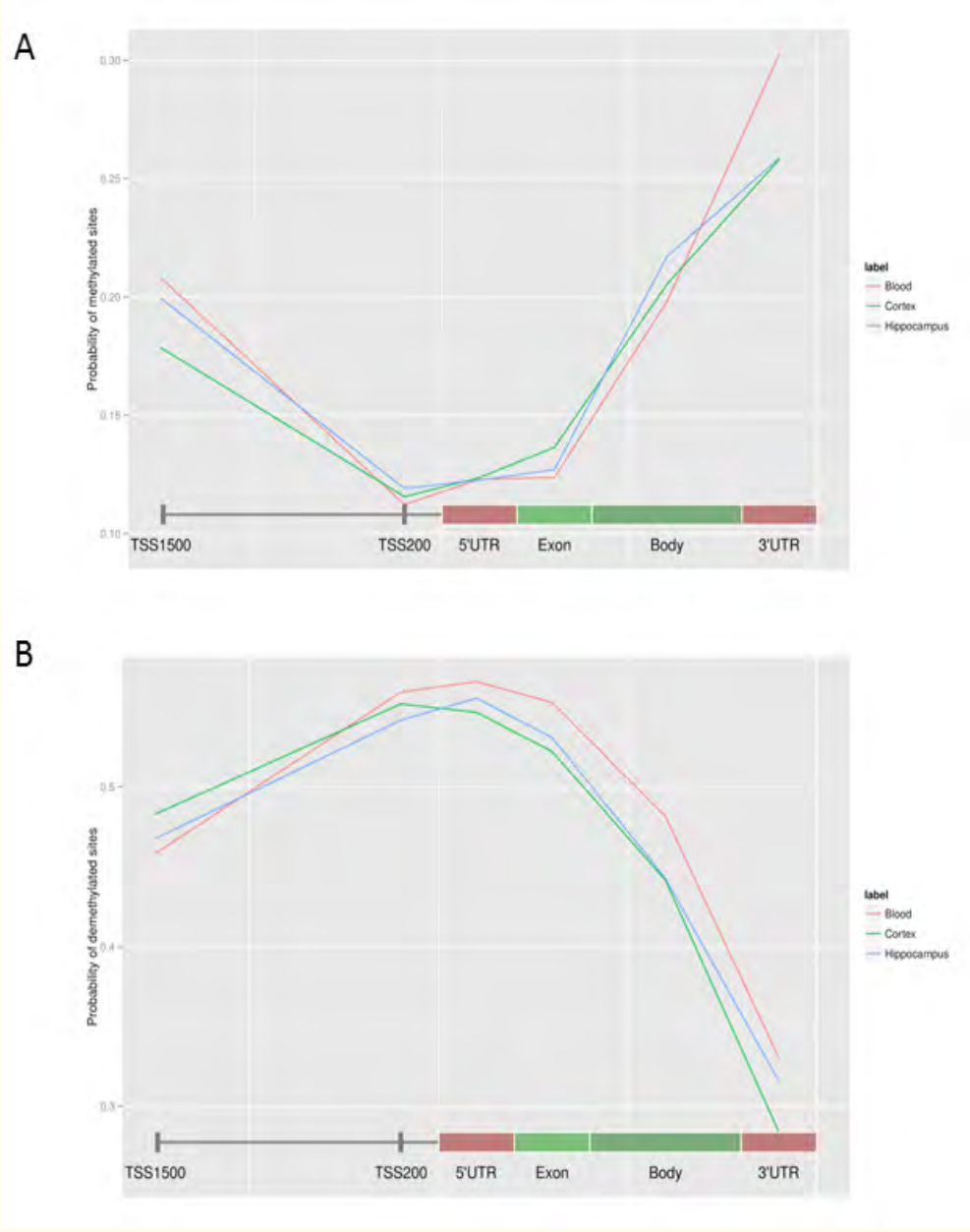


Figure 2: *DNA methylation status across different gene locations. This figure shows a high density of low methylated sites (<0.25) around the start site (TSS200, 1st exon and 5´UTR) and an elevated proportion of medium and high methylated CpG sites in the body and 3´UTR of the gene.*

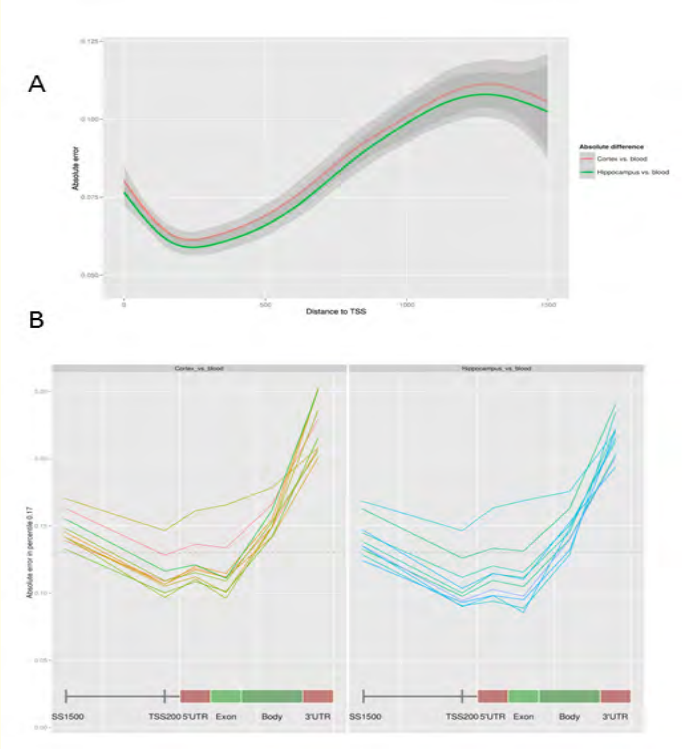


Figure 3: *Distribution of the error across different annotated genomic regions. Subfigure (A) shows the absolute error vs. distance to the TSS. After an initial decrease until approximately 200 bp, the error clearly grows with the distance. Subfigure (B) displays the mean discrete error per site location for each of the samples. Each line represents the mean absolute error made in in those sites located at different genic regions. These plots show that for all the individuals there is a minimum in the error in the proximity of the TSS, particularly in the comparisons between brain and blood.*

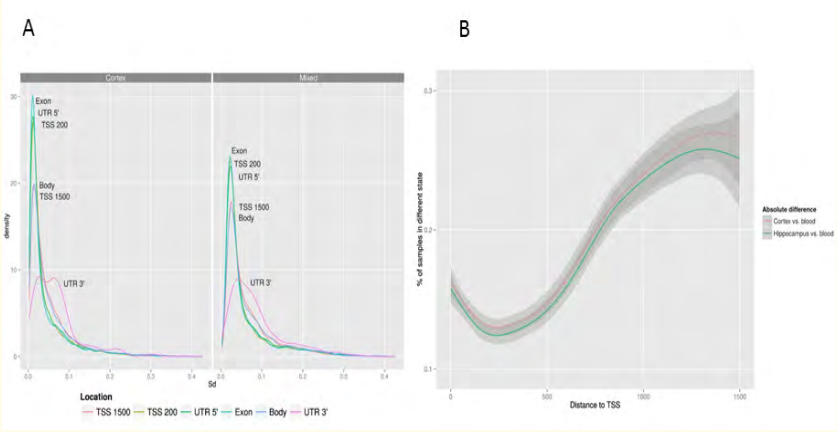


Figure 4: *Distribution of the variability in public cortex data. Subfigure (A) shows the distributions separately for each location. Sites close to the TSS show a lower variability when compared with those far away; (B) shows hos variablity increases as me move away from the TSS.*

Polymorphic CpGs are the ones with the biggest impact on both intra-individual cross-tissue and inter-individual cortex variability

By adapting the list that Chen., *et al*. published in 2011 [31], we determined which of the probes included in the HumanMethylation27K were (1) non-specific and thus with potential cross-reactivity that could compromise true signal detection, (2) target a CpG site that was overlapping a known SNP (i.e., CpGs that are polymorphic at cytosine or guanine positions) or (3) had at least one SNP across the probe. We identified 686 cross-reactive probes (2.5% of the probes), 2,030 CpGs that overlap known SNPs (7.4% of the probes) and 10,700 probes that had a probe that overlapped at least one SNP (38.8% of the probes) (Supplementary Table S2). To determine the extent to which intra-individual cross tissue concordance and inter-individual methylation differences could be affected/confounded by the different types of probes, we compared the distribution of variation among the different probe types (i.e. cross-reactive vs non-cross reactive, polymorphic CpG vs non-polymorphic CpG and with SNPs at probe without SNPs). With respect to the intra-individual cortex-blood variation (measured in error), we found that both polymorphic CpGs and cross-reactive probes had a negative impact on the intra-individual cortex-blood variation (Figure S3). Among the 2,030 probes containing a polymorphic CpG site, only 19.4% displayed low error levels (error < 0.01) while up to 40.6% did so among the non-polymorphic ones (Chi-square test: 355.9, p-value: 2e10-79). As for the cross-reactive probes, we observed that intra-individual variability was also bigger among the cross-reactive probes compared to the non-cross reactive (23.4% vs 39.5% of the probes with intra-individual variation < 0.01, respectively; Chi-square Test: 71.9, p-value: 2x10-17). On the other hand, with regard to the inter-individual cortex variability, as expected, it was bigger among the polymorphic CpGs (19.8% of the sites showing a variability > 0.1), when compared to the non-polymorphic ones (12.0% of the sites showing a variability > 0.1 (Chi-square test: 269.4, p-value: 2x10-60), while cross-reactive probes did not have a significant impact on the inter-individual cortex variation. Overall, our results indicate that while SNPs across the probes do not have a significant impact in any of the variations studied, both cross-reactive probes and polymorphic CpGs can significantly affect to both intra-individual and inter-individual variations.

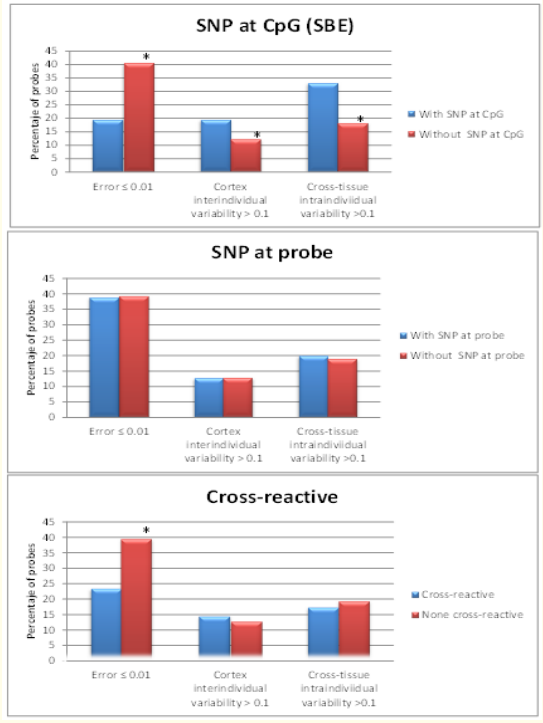
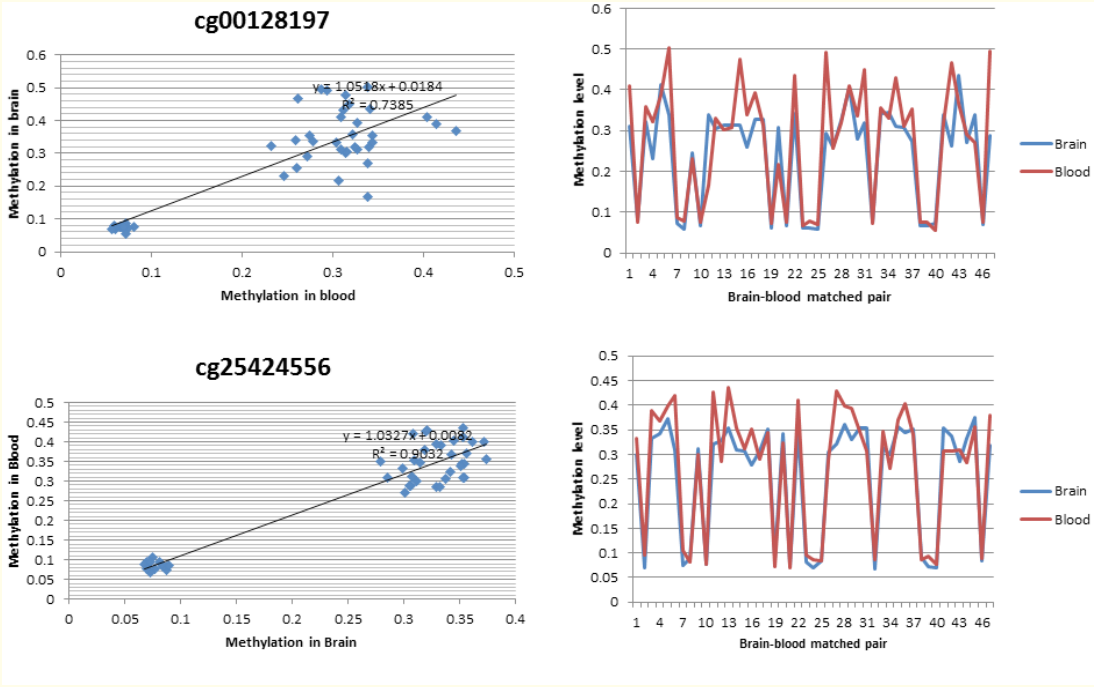


Figure S3: *Effect of the polymorphic CpGs, SNPs at probe or cross-reactive probes on intra-individual and inter-individual variability.*

Validation of the top ranked CpGs in an independent cohort

We have selected the probes showing lowest intra-individual variation (error < 0.01) and filtered out those with a polymorphic CpGs, SNPs at probes and cross-reactive probes by using EpiBrain. Next, based on the matched brain-blood pairs in the GSE59685 publicly available data containing methylation data of 48 matched brain-blood pairs (44 with Alzheimer and 4 healthy controls), we have selected those CpGs that having been classified as the ones with the lowest error by EpiBrain show a relatively high inter-individual difference (SD > 0.1) among the cortex samples in the GSE59685. As it can be seen in Figure S4, which shows the top five CpGs in this analysis, (1) DNA methylation signatures were concordant between blood and brain samples for any of the top ranked 5 CpGs (Mean R2 > 0.85, (2) measurement of DNA methylation in blood would predict 99 % of the outlier samples in brain and (3) these CpGs do not have a different behavior in cases compared to controls, thus further supporting Hannon *et al.*´s findings [14].



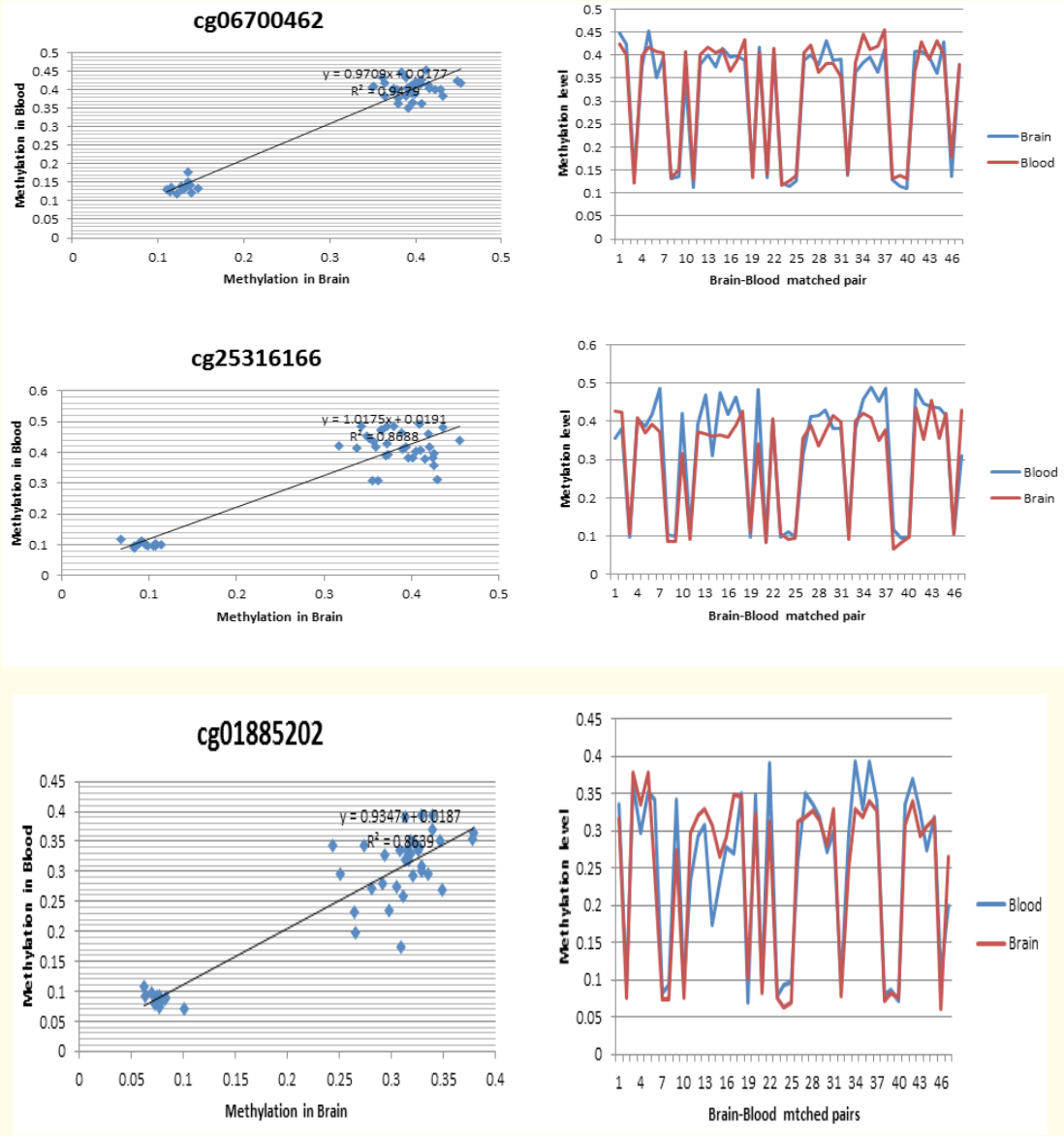


Figure S4: *Replication of two random CpGs showing a low intra-individual variation in an independent set of sample*

Discussion

In this study, we used an array-based approach to interrogate the methylation levels of more than 27,000 CpGs in matched cortex, hippocampus and whole blood samples from ten individuals. The collected data were used to estimate the error one would make when comparing the methylation profile of blood samples to that of brain tissues among others. To the best of our knowledge, there are just a few studies where matched brain-blood samples were used to characterize the intra- and inter-individual methylomic variation across whole blood and brain tissue [13:15], but most of the samples investigated in these studies are from individuals with neurological disease. Our studies in matched blood and brain samples from healthy individuals are consistent with these previous works that Interindividual variation in blood is only correlated across tissues at a small number of sites. Furthermore, the current study provides additional information

such as an inventory of CpG marks for which measurements in the blood can be predictive of the methylation status in the brain, as well as further information concerning the between-individual variability, methylation levels, genomic coordinates, distance to the TSS, non-specific probes or polymorphic CpGs among others.

We have developed an R based software (Epibrain) with the objective of making our results on the intra-individual blood-brain DNA methylation variation available to the scientific community. Epibrain leverages the utility of our paired brain-blood methylation data, by showing the error we would commit if we took blood as a surrogate of the brain (cortex or hippocampus) for any CpG included in the HumanMethylation27K. In order to assess the usefulness of the application, we have selected a set of CpG candidates and validate them in an independent cohort of matched brain-blood pairs in the GSE59685. Additionally, the program also incorporates data from other publicly available methylation data series from GEO in order to show a number of variables such as the DNA methylation inter-individual variability as well as information about the cross-reactive probes or polymorphic CpGs identified by Weksberg´s lab [31] so that the user can easily identify CpGs that can be conditioned by SNPs or cross-reactivity.

As previously reported by Davies., *et al*. [13], we observed that although variation between tissues within individuals is generally greater than variation between individuals, there is a small subset of CpGs showing relatively high between individual variations within the brain tissue and low intra-individual cross-tissue (blood and brain) differences. These CpGs identified in our study could potentially be of great interest for the study of disease-associated molecular alterations taking place in the brain, and some of them have already been shown to play a role in numerous neurodegenerative disorders. For instance, THRSP was found to be deregulated in Alzheimer Disease (AD) and nerve injury [21,22], GRM4´s expression was also altered in AD [23] and it has been reported that knockout mice lacking GMR4 were markedly vulnerable to experimental autoimmune encephalomyelitis [26]. B3GALT1, which is mainly expressed in brain [24,25], was found to be deregulated in an animal model of hypoxic-ischemic encephalopathy. TDG, one more of the aforementioned CpGs and whose aberrant methylation leads to lower gene expression levels [27], codifies for an important enzyme in the base excision repair (BER) pathway. Most of the DNA damage in neurons is repaired by the BER pathway, and it has been reported that aberrant alterations in BER enzymes could be involved in the development of some neurodegenerative diseases such as AD and Parkinson's diseases [28]. Therefore, these lines of evidence suggest that some of these top ranked CpGs (low intra-individual and high inter-individual variation) may be suitable for their application in blood-based epigenetic epidemiologic studies investigating potential risk factors that might contribute to the development of neurodegenerative or psychiatric disorders in human living individuals.

It is interesting to note that when it comes to the distribution of the intra-individual and inter-individuals' variations across different gene locations, both types of variations follow similar patterns (Figure 3 and 4). Certain genetic regions, such as the TSS200, 5´UTR or CGIs, are usually associated with low intra-individual variations as previously reported [29]. But CpGs at these gene locations display in turn very low inter-individual brain differences, and thus they are not in principle good candidates for epidemiologic epigenetic studies. Further analysis of the small subset of 39 probes showing high inter-individual differences within the brain tissue and low intraindividual cross-tissue variations revealed that there was not a significant overrepresentation of any of the gene locations (Figure S2).

Another question of interest is the extents to which marks that show a high epigenetic brain to blood equivalence might also be informative for more internal tissues. Previous studies have reported that although there are some CpGs with differential methylation levels across different somatic tissues, a high proportion of them share similar methylation profiles and the correlation is very high [30]. In this regard, our results indicate that most of the CpG sites showing a low variation between brain and blood also displayed similar methylation levels in a wide range of somatic tissues (SD mixed tissue < 0.1), suggesting that their methylation status in blood would also be informative for other internal tissues.

It has been reported that some of the probes included in both Illumina's 27K and 450K methylation arrays are cross-reactive and/or polymorphic. We note that inter-individual variation was confounded by polymorphic CpG sites, while intra-individual blood-brain variation was affected not only by cross-reactive probes but also by polymorphic CpGs (Figure S3). In this regards, although it was expected that cross-reactive probes could contribute to "artificially" increase the intra-individual DNA methylation variation, this type of variation

was not expected to be affected by the polymorphic CpGs, since these polymorphisms (present in the germ line of an individual) would equally affect both tissues [31]. So far, we are unable to explain how polymorphic CpGs can affect the intra-individual cross-tissue variation, and we believe that further studies should be done to assess the mechanisms by which these polymorphic CpGs could affect the intra-individual cross-tissue variation.

Conclusions

Our results suggest that the error we make when taking a blood sample instead of a brain sample is, in general, greater that the variance we observe between individuals. However, we have identified a small proportion of CpGs for which measurements in blood are representative of methylation in the brain. The identification and analysis of these CpG marks has been done with Epibrain, a newly developed R application.

Conflict of Interest

None to declare.

Authors' Contributions

NGB carried out the molecular genetic studies, participated in the computational analysis and drafted the manuscript. XE and MAA carried out the molecular genetic studies and helped to draft the manuscript. BC performed the computational analysis, created the Epi-Brain tool and drafted the manuscript. II performed the computational analysis and helped to draft the manuscript. AP performed the statistical analysis and helped to draft the manuscript. AA participated in the statistical analysis and helped to draft the manuscript. FGB participated in the design of the study and helped to draft the manuscript. IAA participated in the design of the study and helped to draft the manuscript. RS participated in the design of the study, collected the samples and helped to draft the manuscript. JAL participated in the design of the study, coordinated the computational analysis and helped to draft the manuscript. MMP conceived the study, participated in its design and coordination and helped to draft the manuscript. All authors read and approved the final manuscript.

Acknowledgements

The research is supported by a grant from Department of Health of the Government of the Basque Country (SA-2009/00064). Naiara G. Bediaga is the recipient of postdoctoral fellowships from the Department of Education, Universities, and Research of the Basque Government. Xabier Elcoroaristizabal was the recipient of a Jesús de Gangoiti Barrera Foundation fellowship. The authors are grateful to Ph.D. Maite Alvarez for her technical and human support provided by the DNA Bank Service (SGIker) of the University of the Basque Country (UPV/EHU). The English language in this manuscript has been reviewed by the American Journal Experts (AJE).

Bibliography

1. Ma DK., *et al*. "Epigenetic choreographers of neurogenesis in the adult mammalian brain". *Nature Neuroscience* 13.11 (2010): 1338-1344.

2. Guo JU., *et al*. "Neuronal activity modifies the DNA methylation landscape in the adult brain". *Nature Neuroscience* 14.10 (2011): 1345-1351.

3. Miller CA and Sweatt JD. "Covalent modification of DNA regulates memory formation". *Neuron* 53.6 (2007): 857-869.

4. Renthal W and Nestler EJ. "Histone acetylation in drug addiction". *Seminars in Cell and Developmental Biology* 20.4 (2009): 387-394.

5. Migliore L and Coppede F. "Genetics, environmental factors and the emerging role of epigenetics in neurodegenerative diseases". *Mutation Research* 667.1-2 (2009): 82-97.

6. Kwok JB. "Role of epigenetics in Alzheimer's and Parkinson's disease". *Epigenomics* 2.5 (2010): 671-682.

7. Samaco RC and Neul JL. "Complexities of Rett syndrome and MeCP2". *Journal of Neuroscience* 31.22 (2011): 7951-7959.

8. Mill J., *et al*. "Epigenomic profiling reveals DNA-methylation changes associated with major psychosis". *American Journal of Human Genetics* 82.3 (2008): 696-711.

9. Urdinguio RG., *et al*. "Epigenetic mechanisms in neurological diseases: genes, syndromes, and therapies". *Lancet Neurology* 8.11 (2009): 1056-1072.

10. Madrigano J., *et al*. "Aging and epigenetics: longitudinal changes in gene-specific DNA methylation". *Epigenetics* 7.1 (2012): 63-70.

11. Bind MA., *et al*. "Air pollution and markers of coagulation, inflammation, and endothelial function: associations and epigene-environment interactions in an elderly cohort". *Epidemiology* 23.2 (2012): 332-340.

12. Liu L., *et al*. "Aging, cancer and nutrition: the DNA methylation connection". *Mechanisms of Ageing and Development* 124.10-12 (2003): 989-998.

13. Davies MN., *et al*. "Functional annotation of the human brain methylome identifies tissue-specific epigenetic variation across brain and blood". *Genome Biology* 13.6 (2012): R43.

14. Hannon E., *et al*. "Interindividual methylomic variation across blood, cortex, and cerebellum: implications for epigenetic studies of neurological and neuropsychiatric phenotypes". *Epigenetics* 10.11 (2015): 1024-1032.

15. Walton E., *et al*. "Correspondence of DNA Methylation Between Blood and Brain Tissue and Its Application to Schizophrenia Research". *Schizophr Bulletin* 42.2 (2016): 406-414.

16. Masliah E., *et al*. "Distinctive patterns of DNA methylation associated with Parkinson disease: Identification of concordant epigenetic changes in brain and peripheral blood leukocytes". *Epigenetics* 8.10 (2013): 1030-1038.

17. Provençal N., *et al*. "The signature of maternal rearing in the methylome in rhesus macaque prefrontal cortex and T cells". *Journal of Neuroscience* 32.44 (2012):15626-15642.

18. Ursini G., *et al*. "Stress-related methylation of the catechol-O-methyltransferase Val 158 allele predicts human prefrontal cognition and activity". *Journal of Neuroscience* 31.18 (2011): 6692-6698.

19. Yuferov V., *et al*. "Tissue-specific DNA methylation of the human prodynorphin gene in post-mortem brain tissues and PBMCs". *Pharmacogenetics and Genomics* 21.4 (2011): 185-196.

20. Byun HM., *et al*. "Epigenetic profiling of somatic tissues from human autopsy specimens identifies tissue- and individual-specific DNA methylation patterns". *Human Molecular Genetics* 18.24 (2009): 4808-4817.

21. Swanberg M., *et al*. "Genetically determined susceptibility to neurodegeneration is associated with expression of inflammatory genes". *Neurobiology of Disease* 24.1 (2006): 67-88.

22. Cong L., *et al*. "Genome-wide analysis of DNA methylation in an APP/PS1 mouse model of Alzheimer's disease". *Acta Neurologica Belgica* 114.3 (2014): 195-206.

23. Tan MG., *et al*. "Genome wide profiling of altered gene expression in the neocortex of Alzheimer's disease". *Journal of Neuroscience Research* 88.6 (2010): 1157-1169.

24. Kolbinger F., *et al*. "Cloning of a human UDP-galactose:2-acetamido-2-deoxy-D-glucose 3beta-galactosyltransferase catalyzing the formation of type 1 chains". *Journal of Biological Chemistry* 273.1 (1998): 433-440.

25. Zhou S., *et al*. "[Cloning, sequencing and expression of the full-length gene encoding paramyosin of Schistosoma japonicum in vivo]". *Zhongguo Ji Sheng Chong Xue Yu Ji Sheng Chong Bing Za Zhi* 17.4 (1999): 196-199.

26. Fallarino F., *et al*. "Metabotropic glutamate receptor-4 modulates adaptive immunity and restrains neuroinflammation". *Nature Medicine* 16.8 (2010): 897-902.

27. Peng B., *et al*. "DNA hypermethylation and partial gene silencing of human thymine- DNA glycosylase in multiple myeloma cell lines". *Epigenetics* 1.3 (2006): 138-145.

28. Mantha AK., *et al*. "A short review on the implications of base excision repair pathway for neurons: relevance to neurodegenerative diseases". *Mitochondrion* 16 (2014): 38-49.

29. Bibikova M., *et al*. "High density DNA methylation array with single CpG site resolution". *Genomics* 98.4 (2011): 288-295.

30. Fan S and Zhang X. "CpG island methylation pattern in different human tissues and its correlation with gene expression". *Biochemical and Biophysical Research Communications* 383.4 (2009): 421-425.

31. Chen YA., *et al*. "Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray". *Epigenetics* 8.2 (2013): 203-209.

32. Meyer PE. "R package infotheo Information-theoretic measures 2011.

33. Torfs P and Brauer C. "A (very) short introduction to R". Hydrology and Quantitative Water Management Group (2014).

Volume 5 Issue 3 March 2017