REPORT

Inter-individual methylomic variation across blood, cortex and cerebellum: implications for epigenetic studies of neurological and neuropsychiatric phenotypes.

Eilis Hannon[1], Katie Lunnon[1], Leonard Schalkwyk[2] and Jonathan Mill[1,3*]

[1] University of Exeter Medical School, RILD Building (Level 4), Barrack Road, University of Exeter, Devon, UK.

[2]School of Biological Sciences, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, UK.

[3] Institute of Psychiatry, Psychology & Neuroscience, King's College London, De Crespigny Park, London, UK.

[*] Corresponding author at: University of Exeter Medical School, RILD, Barrack Road, University of Exeter, Devon, UK. UK. Tel: + 44 1392 408 298 Email address: J.Mill@exeter.ac.uk

Eilis Hannon: E.J.Hannon@exeter.ac.uk

Katie Lunnon: K.Lunnon@exeter.ac.uk

Leonard Schalkwyk: lschal@essex.ac.uk

Jonathan Mill: j.mill@exeter.ac.uk

Keywords: DNA methylation, epigenetic epidemiology, brain, cortex, cerebellum, blood, Illumina 450K array.

Abstract:

Given the tissue-specific nature of epigenetic processes, the assessment of disease-relevant tissue is an important consideration for epigenome-wide association studies (EWAS). Little is known about whether easily accessible tissues such as whole blood can be used to address questions about inter-individual epigenomic variation in inaccessible tissues such as the brain. We quantified DNA methylation in matched DNA samples isolated from whole blood and four brain regions (prefrontal cortex, entorhinal cortex, superior temporal gyrus and cerebellum) from 122 individuals. We explored covariation between tissues and the extent to which methylomic variation in blood is predictive of inter-individual variation identified in the brain. For the majority of DNA methylation sites, inter-individual variation in whole blood is not a strong predictor of inter-individual variation in the brain, although the relationship with cortical regions is stronger than with the cerebellum. Variation at a subset of probes is strongly correlated across tissues, even in instances when the actual level of DNA methylation is significantly different between them. A substantial proportion of this covariation, however, is likely to result from genetic influences. Our data suggest that for the majority of the genome, a blood-based EWAS for disorders where brain is presumed to be the primary tissue of interest will give limited information relating to underlying pathological processes. These results do not, however, discount the utility of using a blood-based EWAS to identify biomarkers of disease phenotypyes manifest in the brain. We have generated a searchable database for the interpretation of data from blood-based EWAS analyses (http://epigenetics.iop.kcl.ac.uk/bloodbrain).

Introduction:

There is increasing interest in the role of epigenetic processes in health and disease, with the primary focus of most epigenetic epidemiological studies to date being DNA methylation. [1] Platforms such as the Illumina 450K Human Methylation microarray (450K array) have enabled the economical, high-throughput profiling of methylomic variation across large numbers of samples and epigenome-wide association studies (EWAS), which aim to identify DNA methylation differences associated with environmental exposure and disease, are now underway for many types of pathology including cancer, [2-4] autoimmune disorders, [5, 6] psychiatric phenotypes, [7] neurodevelopmental disorders [8, 9] and dementia. [10, 11] Despite the recent successes in identifying disease-associated epigenetic variation, the design, analysis and interpretation of EWAS requires careful attention; there are a number of critical issues that need to be considered in epigenetic epidemiology that preclude a simple re-analysis of DNA samples collected for genome-wide association studies (GWAS). [9, 12-14]

Of particular importance is the fact that, unlike germline genetic variation, epigenetic signatures are tissue-specific; therefore the selection of tissue type for epigenetic epidemiology is potentially critical. The ENCODE and the NIH Epigenomics Roadmap projects, [15-17] for example, have recently characterized the distinct epigenetic profiles defining human cell-types, highlighting how these reflect the developmental relationships between them. It is clear that intra-individual epigenetic differences (i.e. between tissues within a single person) greatly outweigh inter-individual differences within a specific tissue-type. [18-23] Although many clinical and epidemiological studies are examining epigenetic variation in easily accessible cells obtained from tissues such as whole blood, the extent to which these can be used to address questions about inter-individual epigenomic variation in

inaccessible tissues such as the brain has not yet been systematically explored. Addressing this issue will be critical given the paucity of high-quality brain tissue from clinically well-phenotyped patients and controls, especially if EWAS analyses require sample sizes approaching those necessary to identify genetic associations with complex disease phenotypes. Because the brain and blood originate from different developmental cell lineages and are epigenetically distinct, [23] it is clearly inappropriate to use blood as a proxy measure for *actual* brain DNA methylation profiles. Despite this, epidemiological studies using accessible peripheral tissues may still be informative in an epidemiological context if inter-individual *variation* is correlated across tissues.

In this study we quantified DNA methylation using the 450K array in a unique collection of matched DNA samples isolated from pre-mortem whole blood and four post-mortem brain regions (prefrontal cortex (PFC), entorhinal cortex (EC), superior temporal gyrus (STG) and cerebellum (CER)) dissected at autopsy from 122 individuals. We describe patterns of covariation across tissues and identify sites where estimates of DNA methylation in whole blood are predictive of inter-individual variation in DNA methylation across the four brain regions. Our data are available in an online searchable database (http://epigenetics.iop.kcl.ac.uk/bloodbrain) to enable the research community to explore the relationship between whole blood and brain DNA methylation patterns at specific locations across the genome.

Results and Discussion:

Cortex, cerebellum and blood are defined by very distinct profiles of DNA methylation

We used the Illumina 450K array to quantify DNA methylation in four dissected brain regions (PFC: n = 114, EC: n = 108, STG: n = 117 and CER: n = 112) and matched pre-

mortem whole blood samples (n = 80) from an overlapping set of 122 individuals (**Supplementary Table 1**). Following pre-processing, normalization and stringent quality control (see **Materials and Methods**), principal component (PC) analysis was performed across the full dataset (comprising 531 individual DNA samples and data for a pruned set of 427,018 high-quality probes). The first PC, explaining 51.4% of the variance, clearly distinguishes between whole blood, cerebellum, and the three cortical regions (**Supplementary Figure 1**). A similar separation of these tissues is observed with the second PC which explains 29.4% of the variance. This observation concurs with previous studies comparing whole blood and brain samples based on smaller samples and using alternative technologies for assessing DNA methylation. [23-25] These differences between tissues are reflected in gene expression data, [26, 27] which confirms the cerebellum as being clearly distinguishable from cortical brain regions. The three cortical regions have strikingly similar DNA methylation profiles; although examination of further PCs, none of which explain more than 5% of the variance, starts to tease apart these regions, none of the top 20 PCs does this definitively (**Supplementary Figure 1**).

Inter-individual variation and sex make a much smaller contribution to overall variation in autosomal DNA methylation than tissue differences

We used linear regression models to calculate the proportion of variance in DNA methylation explained by tissue, individual and sex. Across autosomal probes, tissue is the strongest predictor of DNA methylation (**Supplementary Figure 2**), although there is a subset of probes for which individual predicts more of the variance (5.39% of all probes) than tissue type (see **Supplementary Table 2**). Across all autosomal probes passing stringent QC (n = 416,872), tissue explains > 50% of the variance in DNA methylation at 193,333 (46.4%) sites, compared to individual differences which explain > 50% of the variance at 4,669

(1.12%) sites (**Figure 1A**). These percentages increase to 66.2% and 1.61% respectively when limited to probes classified as being "blood variable" (n = 185,060) (see **Materials and Methods)**. As expected, sex makes a strong contribution to variation observed at probes on chromosomes X and Y (n = 10,146), explaining >50% of the variance at 5,920 (58.3%) of these positions compared to tissue which explains >50% of the variance at only 1,359 (1.34%) sites. In contrast, sex makes a very small contribution to autosomal variation, explaining >50% of the variance at only 18 ($4.32x10^{-3}$%) of autosomal probes (**Supplementary Figure 3**).

Inter-individual variation is correlated across tissues at a small number of sites

Although patterns of DNA methylation are clearly distinct across different brain regions and blood, driven by highly significant mean differences in DNA methylation at multiple sites across the genome, [23] we were interested in exploring the extent to which inter-individual variation detected in blood reflects inter-individual variation in the three cortical regions and cerebellum. Focusing on autosomal DNA methylation sites characterized as being variable in whole blood (n = 185,060, see **Materials and Methods**), correlation coefficients across all individuals were calculated between DNA methylation in whole blood and each of the four brain regions. Randomly selecting pairs of 'matched' samples to derive the null distribution, we found a modest but highly-significant positive shift in the distribution of correlations for each of the four brain regions (PFC: Wilcoxon test $P < 2.2x10^{-308}$, EC: $P < 2.2x10^{-308}$, STG: $P < 2.2x10^{-308}$, CER: $P < 2.2x10^{-308}$), with a small peak highlighting a number of probes characterized by a near perfect correlation (**Figure 1B-E).** For the majority of probes, however, inter-individual variation in DNA methylation in whole blood explains only a small amount of the variation seen in any of the brain regions (**Figure 2**). For example, DNA methylation in whole blood is strongly correlated with levels in cerebellum (i.e. explaining

>50% of the variance) for only 1.19% of "blood variable" probes, and moderately correlated (i.e. explaining >20% of the variance) with 3.68% of "blood variable" probes. Of note, the extent of inter-individual correlation is significantly higher ($P < 1.0^{-308}$) between whole blood and each of the three cortical regions than with the cerebellum, although the proportion of correlated probes is still low (**Supplementary Table 3**). These data concur with previous small studies correlating inter-individual variation in DNA methylation between tissues. Slieker *et al.* compared DNA methylation profiles between blood and several internal organs and reported a comparable number of sites (5,532, 3,909, 10,905, and 2,446 sites for liver, subcutaneous fat, omentum and skeletal muscle respectively) with a strong relationship ($r > 0.8$) with variation in blood. [19] Similarly, a study comparing DNA methylation in matched blood and buccal samples found only ~3% of 998 sites were characterized by an absolute Pearson correlation > 0.5. [20] Density plots of DNA methylation across the probes with the strongest positive correlations (> 0.95) between blood and brain indicate that many are characterized by a clear trimodal distribution of DNA methylation **(Supplementary Figure 4**), suggesting that DNA sequence variation likely mediates much of the observed cross-tissue similarities via processes such as allele-specific DNA methylation. [28] Many DNA methylation quantitative trait loci (mQTL) have consistent effects across tissues [29, 30]; **Supplementary Figure 5** shows a couple of examples where the correlated DNA methylation profiles across tissues is likely to result from mQTLs, as the distribution of DNA methylation levels cluster into distinct groups reflecting genotype, with consistent effects across tissues.

For some loci, the extent to which inter-individual variation is correlated between whole blood and brain differs by brain region

The extent to which inter-individual variation in whole blood is correlated with that in the brain is similar across the four brain regions (**Supplementary Figure 6**), with a highly significant correlation of probe-wise correlations. Of note, the three cortical regions are more similar to each other in this regard than the cerebellum, indicating that there is a subset of probes where variation in whole blood predicts variation in the cortex and not the cerebellum, and vice versa (for example see **Supplementary Figure 7**). This suggests that the extent of covariation between pairs of tissues can differ depending upon the tissues in question, and establishing a correlation between any two tissues does not imply a correlation between all tissues. Of additional interest are sites where there is a significant, but *negative*, correlation between blood and brain (see **Supplementary Figure 8** for specific examples), a phenomenon that has been reported previously for certain loci. [25] This phenomenon is relatively rare, and most notable between DNA methylation in whole blood and cerebellum (**Supplementary Table 4**).

Sites at which inter-individual variation in DNA methylation is highly correlated between whole blood and brain are enriched in CpG-rich promoter regions

Sites at which DNA methylation is strongly correlated between whole blood and brain ($r^2 > 0.5$) are not equally distributed across the genome. Of note, we find a significant over-representation at loci in the vicinity of transcription start sites (PFC: $P = 1.34x10^{-22}$, EC: $P = 5.15x10^{-21}$, STG: $P = 1.06x10^{-18}$, CER: $P = 1.34x10^{-22}$), 1st exon (PFC: $P = 2.48 \times 10^{-175}$, EC: $P = 8.78x10^{-174}$, STG: $P = 6.72x10^{-171}$, CER: $P = 3.50x10^{-172}$) and 5'UTR (PFC: $P = 2.07 \times 10^{-119}$, EC: $P = 5.04 \times 10^{-120}$, STG: $P = 2.98x10^{-119}$, CER: $P = 2.28x10^{-118}$) and a depletion in the gene body (PFC: $P = 6.15x10^{-94}$, EC: $P = 4.43x10^{-97}$, STG: $P = 2.56x10^{-87}$, CER: $P = 1.85x10^{-88}$), 3'UTR (PFC: $P = 2.92x10^{-24}$, EC: $P = 1.06 \times 10^{-23}$, STG: $P = 2.02 \times 10^{-21}$, CER: $P = 4.74x10^{-23}$) and intergenic regions (PFC: $P = 1.10x10^{-72}$, EC: $P = 2.24 \times 10^{-65}$, STG: P =

3.51 x 10^{-71}, CER: P = 1.57 x 10^{-86}) (**Figure 4** and **Supplementary Table 5**). In addition there is enrichment in CpG islands (PFC: P < $2.2x10^{-308}$, EC: P < $2.2x10^{-308}$, STG: P < $2.2x10^{-308}$, CER: P < $2.2x10^{-308}$) and depletion in open sea (PFC: P = $8.16x10^{-262}$, EC: P = 3.57 x 10^{-276}, STG: P = $2.87x10^{-273}$, CER: P = $7.10x10^{-244}$) (**Figure 4** and **Supplementary Table 5**).

Covariation between tissues often occurs when absolute levels of DNA methylation are different

It is a common misconception that a similar average level of DNA methylation between two tissues at a given locus is sufficient to establish that one of these tissues may be used as a proxy for the other. [31] In fact, for epidemiological studies that use peripheral tissues as a proxy, it is actually more important that the two tissues *covary*, regardless of their absolute DNA methylation levels. To demonstrate this point, **Supplementary Table 6** lists 887 sites characterized by similar levels of DNA methylation between tissues (paired t-test P > 0.1) but no evidence for inter-individual covariation (r^2 < 0.05), with specific examples shown in **Supplementary Figure 9**. In contrast, **Supplementary Table 7** and **Supplementary Figure 10** demonstrate sites that are characterized by highly tissue-specific levels of DNA methylation (paired t-test P < 0.00001) but strong evidence for inter-individual covariation (r^2 > 0.5).

Whole blood cannot be used as a proxy for DNA methylation sites that are only variable in the brain

One potential caveat to performing an EWAS of a neurological/psychiatric phenotype using a peripheral tissue as a proxy is that a proportion of sites are characterized by limited inter-individual variation in whole blood but high levels of inter-individual variation in the brain,

and vice versa. We defined probes as having 'low' variation when the range of DNA methylation values across the total sample < 5% and 'high' variation when the range of DNA methylation in the middle 80th percentile of samples > 5%. **Figure 5A** shows that there are 2,505 sites characterized by high inter-individual variation in whole blood but not in the cortex (STG) (see **Supplementary Figure 11** and **Supplementary Table 8** for corresponding data for the other cortical regions) and 6,909 sites that vary in whole blood but not in the cerebellum. Whether these sites are omitted from analyses (for example to reduce multiple testing burden) depends upon the ultimate aim of the study being undertaken; while they may not be able to inform directly about mechanistic processes in disease they could still represent useful biomarkers. In contrast, some sites are non-variable in whole blood but vary in the brain (see **Figure 5B)**.

Conclusion:

Our data suggest that across the majority of the genome, an EWAS using whole blood for disorders where brain is the presumed to be the primary tissue of interest will give limited information relating to underlying pathological processes. However, there are a proportion of sites where inter-individual variation is correlated between whole blood and brain, and these results do not discount the utility of using a blood-based EWAS to identify potential biomarkers of psychiatric disease phenotypes. We have developed a searcheable online database (http://epigenetics.iop.kcl.ac.uk/bloodbrain/) to enable researchers to investigate the relationship between blood and brain for any probes on the Illumina 450K Beadchip array to aid in the interpretation of EWAS analyses of brain disorders.

Materials and Methods:

***Samples*.**

We obtained entorhinal cortex (EC), prefrontal cortex (PFC), superior temporal gyrus (STG), and cerebellum (CER) tissue from 117 individuals archived in the MRC London Neurodegenerative Disease Brain Bank (http://www.kcl.ac.uk/iop/depts/cn/research/MRC-London-Neurodegenerative-Diseases-Brain-Bank/MRC-London-Neurodegenerative-Diseases-Brain-Bank.aspx). Ethical approval for the study was provided by the NHS South East London REC 3. All samples were dissected by trained specialists, snap-frozen and stored at −80 °C. Matched whole blood samples collected before death were available for 80 samples (**Supplementary Table 1**) as part of the Alzheimer's Research UK funded study "Biomarkers of AD Neurodegeneration", with informed consent according to the Declaration of Helsinki (1991). Genomic DNA was isolated from ~100 mg of each dissected brain region or ~10ml whole blood stored in EDTA collection tubes using a standard phenol-chloroform extraction method, and tested for degradation and purity before analysis.

Methylomic profiling

500ng DNA from each sample was sodium bisulfite-treated using the Zymo EZ 96 DNA methylation kit (Zymo Research) according to the manufacturer's standard protocol. DNA methylation was quantified using the Illumina Infinium Human Methylation450K BeadChip (Illumina) using a Illumina HiScan System (Illumina). All samples were assigned a unique code for the purpose of the experiment and grouped by tissue and randomized with respect to other variables status to avoid batch effects, and processed in batches of four BeadChips. Illumina Genome Studio software was used to extract the raw signal intensities of each probe (without background correction or normalization). Raw data are downloadable from GEO with accession identifier GSE59685.

Data pre-processing

All analyses were performed using R 3.0.2 [32] and Bioconductor 2.13. [33] Signal intensities were imported into R using the *methylumi* package. [34] Initial quality control checks were performed using functions in the *methylumi* package to assess concordance between reported and gender inferred from the arrays. The 65 non-CpG SNP probes on the array were also used to confirm that all four brain regions and matched blood samples were sourced from the same individual. Data was subsequently pre-processed in the R package *wateRmelon* using the *dasen* function as previously described. [35] Prior to data analysis, we removed the 65 non-CpG SNP probes, and probes characterized by either non-specific binding or containing common (minor allele frequency > 5%) SNPs within 10bp of the CG or single base extension position, identified from previously published lists. [36, 37]

Data analysis

Separate linear regression models were used to calculate the proportion of variance explained (adjusted r squared) by a) brain region, b) individual, and c) sex for each DNA methylation site on the array across individuals for which data from all five tissues passed quality control. A subset of "blood variable" probes was identified by calculating the DNA methylation difference between the 10^{th} and 90^{th} percentile across all samples, and selecting sites where this was > 5% (all chromosomes n = 194,426; autosomes n = 185,060). Sites characterized by overall differential DNA methylation between blood and each brain region were identified by a paired t-test of matched samples. Pairwise correlation coefficients were calculated between DNA methylation values from whole blood and each of the four brain regions across matched samples from linear regression models; the values were squared and multiplied by 100 to obtain the percentage of variance explained for each probe. Samples were permuted

and correlations between DNA methylation in whole blood and brain were recalculated across unmatched pairs to establish the distribution in the scenario where there is no relationship between DNA methylation in blood and brain. The density curve of these simulated correlations was added to the histograms of the true correlation coefficients to represent the null distribution (**Figure 1** and **Supplementary Figure 4**). The annotation file provided by Illumina for all probes on the array was used to classify DNA methylation sites into genomic feature and CpG island feature categories; any site with no UCSC gene annotation was classed as "intergenic". Enrichment was calculated from a 2x2 Fisher's Exact test, comparing the number of probes with blood-brain correlation $r^2 > 0.5$ annotated to each feature category to the background of all probes.

Web resources.

A searchable database of matched blood and brain region DNA methylation data is available at http://epigenetics.iop.kcl.ac.uk/bloodbrain/. It reports the distribution of DNA methylation values in each tissue and the correlation of individual values between blood and each of the four brain regions for each probe on the 450K array.

Acknowledgments

We thank the National Institute for Health (NIHR) Biomedical Research Unit in Dementia in the South London and Maudsley NHS Foundation Trust (SLaM), Brains for Dementia Research (Alzheimer Brain Bank, UK), and the donors and families who made this research possible. Blood samples were collected as part of the Alzheimer's Research UK funded study "Biomarkers of AD Neurodegeneration". This work was funded by US National Institutes of Health grant R01 AG036039 and UK Medical Research Council grant MR/K013807/1 to J.M.

References

1. Murphy TM, Mill J. Epigenetics in health and disease: heralding the EWAS era. Lancet 2014; 383:1952-4.
2. Heyn H, Carmona FJ, Gomez A, Ferreira HJ, Bell JT, Sayols S, et al. DNA methylation profiling in breast cancer discordant identical twins identifies DOK7 as novel epigenetic biomarker. Carcinogenesis 2013; 34:102-8.
3. Irizarry RA, Ladd-Acosta C, Wen B, Wu Z, Montano C, Onyango P, et al. The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. Nat Genet 2009; 41:178-86.
4. Lange CP, Campan M, Hinoue T, Schmitz RF, van der Meulen-de Jong AE, Slingerland H, et al. Genome-scale discovery of DNA-methylation biomarkers for blood-based detection of colorectal cancer. PLoS One 2012; 7:e50266.
5. Liu Y, Aryee MJ, Padyukov L, Fallin MD, Hesselberg E, Runarsson A, et al. Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. Nat Biotechnol 2013; 31:142-7.
6. Rakyan VK, Beyan H, Down TA, Hawa MI, Maslau S, Aden D, et al. Identification of type 1 diabetes-associated DNA methylation variable positions that precede disease diagnosis. Plos Genet 2011; 7:e1002300.
7. Pidsley R, Viana J, Hannon E, Spiers HH, Troakes C, Al-Saraj S, et al. Methylomic profiling of human brain tissue supports a neurodevelopmental origin for schizophrenia. Genome Biol 2014; 15:483.
8. Ladd-Acosta C, Hansen K, Briem E, Fallin M, Kaufmann W, Feinberg A. Common DNA methylation alterations in multiple brain regions in autism. Molecular Psychiatry 2014; 19:862-71.
9. Berko ER, Suzuki M, Beren F, Lemetre C, Alaimo CM, Calder RB, et al. Mosaic epigenetic dysregulation of ectodermal cells in autism spectrum disorder. Plos Genet 2014; 10:e1004402.
10. De Jager PL, Srivastava G, Lunnon K, Burgess J, Schalkwyk LC, Yu L, et al. Alzheimer's disease: early alterations in brain DNA methylation at ANK1, BIN1, RHBDF2 and other loci. Nat Neurosci 2014; 17:1156-63.
11. Lunnon K, Smith R, Hannon E, De Jager PL, Srivastava G, Volta M, et al. Methylomic profiling implicates cortical deregulation of ANK1 in Alzheimer's disease. Nat Neurosci 2014; 17:1164-70.
12. Mill J, Heijmans BT. From promises to practical strategies in epigenetic epidemiology. Nat Rev Genet 2013; 14:585-94.
13. Relton CL, Davey Smith G. Epigenetic epidemiology of common complex disease: prospects for prediction, prevention, and treatment. PLoS Med 2010; 7:e1000356.
14. Rakyan VK, Down TA, Balding DJ, Beck S. Epigenome-wide association studies for common human diseases. Nat Rev Genet 2011; 12:529-41.
15. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature 2012; 489:57-74.
16. Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, et al. Mapping and analysis of chromatin state dynamics in nine human cell types. Nature 2011; 473:43-9.
17. Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, et al. Integrative analysis of 111 reference human epigenomes. Nature 2015; 518:317-30.

18. Byun HM, Siegmund KD, Pan F, Weisenberger DJ, Kanel G, Laird PW, et al. Epigenetic profiling of somatic tissues from human autopsy specimens identifies tissue- and individual-specific DNA methylation patterns. Hum Mol Genet 2009; 18:4808-17.
19. Slieker RC, Bos SD, Goeman JJ, Bovée JV, Talens RP, van der Breggen R, et al. Identification and systematic annotation of tissue-specific differentially methylated regions using the Illumina 450k array. Epigenetics Chromatin 2013; 6:26.
20. Jiang R, Jones MJ, Chen E, Neumann SM, Fraser HB, Miller GE, et al. Discordance of DNA methylation variance between two accessible human tissues. Sci Rep 2015; 5:8257.
21. Thompson TM, Sharfi D, Lee M, Yrigollen CM, Naumova OY, Grigorenko EL. Comparison of whole-genome DNA methylation patterns in whole blood, saliva, and lymphoblastoid cell lines. Behav Genet 2013; 43:168-76.
22. Eckhardt F, Lewin J, Cortese R, Rakyan VK, Attwood J, Burger M, et al. DNA methylation profiling of human chromosomes 6, 20 and 22. Nat Genet 2006; 38:1378-85.
23. Davies MN, Volta M, Pidsley R, Lunnon K, Dixit A, Lovestone S, et al. Functional annotation of the human brain methylome identifies tissue-specific epigenetic variation across brain and blood. Genome Biol 2012; 13:R43.
24. Farré P, Jones MJ, Meaney MJ, Emberly E, Turecki G, Kobor MS. Concordant and discordant DNA methylation signatures of aging in human blood and brain. Epigenetics Chromatin 2015; 8:19.
25. Illingworth RS, Gruenewald-Schneider U, De Sousa D, Webb S, Merusi C, Kerr AR, et al. Inter-individual variability contrasts with regional homogeneity in the human brain DNA methylome. Nucleic Acids Res 2015; 43:732-44.
26. GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science 2015; 348:648-60.
27. Melé M, Ferreira PG, Reverter F, DeLuca DS, Monlong J, Sammeth M, et al. Human genomics. The human transcriptome across tissues and individuals. Science 2015; 348:660-5.
28. Schalkwyk LC, Meaburn EL, Smith R, Dempster EL, Jeffries AR, Davies MN, et al. Allelic skewing of DNA methylation is widespread across the genome. Am J Hum Genet 2010; 86:196-212.
29. Gibbs JR, van der Brug MP, Hernandez DG, Traynor BJ, Nalls MA, Lai SL, et al. Abundant Quantitative Trait Loci Exist for DNA Methylation and Gene Expression in Human Brain. Plos Genet 2010; 6.
30. Smith AK, Kilaru V, Kocak M, Almli LM, Mercer KB, Ressler KJ, et al. Methylation quantitative trait loci (meQTLs) are consistently detected across ancestry, developmental stage, and tissue type. Bmc Genomics 2014; 15:145.
31. Smith AK, Kilaru V, Klengel T, Mercer KB, Bradley B, Conneely KN, et al. DNA extracted from saliva for methylation studies of psychiatric traits: evidence tissue specificity and relatedness to brain. Am J Med Genet B Neuropsychiatr Genet 2015; 168B:36-44.
32. R Development Core Team. R: A Language and Environment for Statistical Computing. . Vienna, Austria: R Foundation for Statistical Computing, 2008.
33. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, et al. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol 2004; 5:R80.
34. Davis S, Du P, Bilke S, Triche J, Bootwalla M. *methylumi: Handle Illumina methylation data. R package version 2.14.0.*, 2015.
35. Pidsley R, Y Wong CC, Volta M, Lunnon K, Mill J, Schalkwyk LC. A data-driven approach to preprocessing Illumina 450K methylation array data. Bmc Genomics 2013; 14:293.

36. Chen YA, Lemire M, Choufani S, Butcher DT, Grafodatskaya D, Zanke BW, et al. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. Epigenetics 2013; 8:203-9.
37. Price ME, Cotton AM, Lam LL, Farré P, Emberly E, Brown CJ, et al. Additional annotation enhances potential for biologically-relevant analysis of the Illumina Infinium HumanMethylation450 BeadChip array. Epigenetics Chromatin 2013; 6:4.

Figure Legends

Figure 1: Variation in DNA methylation in whole blood is correlated with variation in the brain for a small proportion of probes. A) the proportion of sites (y-axis) for which tissue (black), sex (red), or individual (green) explain a given percentage of DNA methylation variance (x-axis). B) to E) histograms showing the distribution of correlation coefficients between DNA methylation in whole blood and the four brain regions (PFC, EC, STG and CER). For all four brain regions the distribution of correlation coefficients is significantly skewed to the right, with stronger correlations seen between whole blood and cortical regions than between whole blood and cerebellum.

Figure 2: Variation in DNA methylation in whole blood as a predictor of variation in the brain. Shown is the proportion of sites (y-axis) for which variation in blood explains a certain of percentage of DNA methylation variance (x-axis) in the PFC (black), EC (red), STG (green) and CER (blue) from the same individuals.

Figure 3: DNA methylation in whole blood significantly covaries with that in the brain at some genomic loci. An example output of our online database (http://epigenetics.iop.kcl.ac.uk/bloodbrain) for blood-brain correlations at cg26039926. Shown is a boxplot of the distribution of DNA methylation values across all individuals split by tissue and four scatterplots demonstrating the relationship between DNA methylation in whole blood and four brain regions (PFC, EC, STG, CER). At this probe there is a highly significant correlation between individual variation in whole blood and that observed in all four brain regions.

Figure 4: Sites at which inter-individual variation correlates between whole blood and brain are enriched in specific genic features. Bar charts plotting the percentage of sites annotated to particular genic feature categories and CpG Island annotations for the full set of "blood variable" sites, in addition to the subset of sites characterized by the highest correlation ($r^2 > 50\%$) between blood and brain. Fisher's exact tests were used to test for either over or underrepresentation for each type of feature and are presented in **Supplementary Table 2**.

Figure 5: EWAS analyses of brain phenotypes using whole blood DNA may potentially miss disease associated variation and interrogate DNA methylation sites that are not actually variable in the brain. Venn diagrams showing the overlap of DNA methylation sites that are A) variable in whole blood but not variable in the cortex (STG) or cerebellum and B) variable in the cortex (STG) and cerebellum but not in whole blood.

Supplementary Tables

Supplementary Table 1: Summary of demographic information for all samples included in this study.

Supplementary Table 2: Top 100 autosomal sites where individual predicts more of the variance in DNA methylation than tissue. Full table can be found at http://epigenetics.iop.kcl.ac.uk/bloodbrain/SupplementaryTables.xlsx

Supplementary Table 3: Counts of sites where DNA methylation in blood is correlated with brain.

Supplementary Table 4: Counts of sites where DNA methylation in blood is negatively correlated with brain.

Supplementary Table 5: Results of tests for enrichment of blood-brain correlated DNA methylation sites in specific genomic or CpG island annotation categories.

Supplementary Table 6: Table of 100 example sites characterized by similar levels of DNA methylation across blood and brain (paired t-test $P > 0.1$) but no evidence of inter-individual covariation ($r^2 < 0.05$). The full table of 887 sites can be found at http://epigenetics.iop.kcl.ac.uk/bloodbrain/SupplementaryTables.xlsx

Supplementary Table 7: Table of 100 example sites characterized by highly tissue-specific levels of DNA methylation (paired t-test $P < 0.00001$) but strong evidence for inter-individual covariation ($r^2 > 0.5$). The full table of 1,813 sitescan be found at http://epigenetics.iop.kcl.ac.uk/bloodbrain/SupplementaryTables.xlsx

Supplementary Table 8: Table of 100 example sites which do not vary in whole blood but do vary in at least one brain region. The full table of 7,821 sites can be found at http://epigenetics.iop.kcl.ac.uk/bloodbrain/SupplementaryTables.xlsx