Whole Genome Grey and White matter DNA Methylation Profiles in Dorsolateral Prefrontal cortex

Jose Vicente Sanchez-Mut, Holger Heyn, Enrique Vidal, Raúl Delgado-Morales, Sebastian Moran, Sergi Sayols, Juan Sandoval, Isidre Ferrer, Manel Esteller and Johannes Gräff

1 Brain Mind Institute, Faculty of Life Sciences, Ecole Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland
2 Single Cell Genomics Unit, Centre Nacional d'Anàlisi Genòmica, E-08028 Barcelona, Catalonia, Spain.
3 Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Barcelona E-08003, Spain; Universitat Pompeu Fabra (UPF), Barcelona, Spain.
4 Cancer Epigenetics Group, Cancer Epigenetics and Biology Program, Institut d'Investigació Biomèdica de Bellvitge, E-08908 Barcelona, Spain.
5 Department of Psychiatry and Neuropsychology, School for Mental Health and Neuroscience (MHeNS), Maastricht University, Universiteitsring 50, 6200 MD, Maastricht, The Netherlands.
6 Bioinformatics Core Facility, Institute of Molecular Biology, D-55128 Mainz, Germany.
7 Biomarkers and Precision Medicine Unit, Health Research Institute La Fe, Valencia 46026, Spain.
8 Institute of Neuropathology, Bellvitge University Hospital, University of Barcelona, Centre for Biomedical Research on Neurodegenerative Diseases (CIBERNED), E-08908 Hospital de Llobregat, Spain.

* Equal contribution

Running title: Grey and White matter methylomes

To whom correspondence should be addressed:

Johannes Gräff, École Polytechnique Fédérale de Lausanne (EPFL), Sciences de la vie (SV), Brain Mind Instute (BMI), Station 19, CH-1015 Lausanne, Switzerland. Tel: +41 21 69 30713. Fax: +41 21 69 30955, johannes.graeff@epfl.ch

Manel Esteller, Programa d'Epigenètica i Biologia del Càncer (PEBC), Fundació Institut d'Investigació Biomèdica de Bellvitge (IDIBELL), Av. Gran Via de l'Hospitalet, 199-203, E-08908 L'Hospitala del Llobregat (Barcelona), Spain. Tel: +34 93 260 71 21. Fax: +34 93 260 72 19. mesteller@idibell.cat

Keywords: DNA methylation, Alzheimer’s disease, Parkinson’s disease, Multiple sclerosis, Amyotrophic lateral sclerosis

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an ‘Accepted Article’, doi: 10.1002/syn.21959

© 2017 Wiley Periodicals, Inc.

Received: Sep 30, 2016; Revised: Jan 10, 2017; Accepted: Jan 11, 2017

This article is protected by copyright. All rights reserved.
Abstract

The brain’s neocortex is anatomically organized into grey and white matter, which are mainly composed by neuronal and glial cells, respectively. The neocortex can be further divided in different Brodmann areas according to their cytoarchitectural organization, which are associated with distinct cortical functions. There is increasing evidence that brain development and function are governed by epigenetic processes, yet their contribution to the functional organization of the neocortex remains incompletely understood. Herein, we determined the DNA methylation patterns of grey and white matter of dorsolateral prefrontal cortex (Brodmann area 9), an important region for higher cognitive skills that is particularly affected in various neurological diseases. For avoiding inter-individual differences, we analyzed white and grey matter from the same donor using whole genome bisulfite sequencing, and for validating their biological significance, we used Infinium HumanMethylation450 BeadChip and pyrosequencing in ten and twenty independent samples respectively. The combination of these analysis indicated robust grey-white matter differences in DNA methylation. What is more, cell type-specific markers were enriched among the most differentially methylated genes. Interestingly, we also found an outstanding number of grey-white matter differentially methylated genes that have previously been associated with Alzheimer’s, Parkinson’s, and Huntington’s disease, as well as Multiple and Amyotrophic lateral sclerosis. The data presented here thus constitute an important resource for future studies not only to gain insight into brain regional as well as grey and white matter differences, but also to unmask epigenetic alterations that might underlie neurological and neurodegenerative diseases.
Introduction

The neocortex is the largest structure in the human brain representing 82% of total brain mass and containing more than 80 billion cells. It is the brain region where cognition, language and complex behaviors take place and are being planned, and where neurons and glial cells form a patterned structure that defines the two main divisions of the brain: the grey and the white matter. The grey matter, the greyish appearance of which is the consequence of the presence of cell bodies, is mainly composed by neurons and glial cells (glia/neuron ratio ~ 1.5) whereas the white matter, the whiteness of which is due to the myelin that covers axons, is mostly composed by glial cells (glia/neuron ratio ~ 15.5) (Azevedo et al., 2009). Therefore, unraveling the epigenetic profiles governing grey and white matter will provide information about the epigenetic landscapes of neurons and glial cells. This quest has recently started to be addressed using genome-wide (Guintivano et al., 2013; Iwamoto et al., 2011; Kessler et al., 2016; Kozlenkov et al., 2014; Labonte et al., 2013) and whole-genome approaches (Lister et al., 2013). In these studies, immunolabeled neuron and glial cells were isolated and their DNA methylation patterns examined allowing for the comparison of pure populations of cellular types, yet without considering their spatial organization. This is an important handicap because several studies have highlighted the importance of considering anatomical compartments. Indeed, neuron and glial cells show specific spatial profiles of gene expression according to their emplacement within anatomical structures (Bohland et al., 2010; Ko et al., 2013), which is also evident by characteristic DNA methylation profiles indicating an important spatial regulation as shown in other brain areas (Davies et al., 2012; Hernandez et al., 2011; Kozlenkov et al., 2016; Ladd-Acosta et al., 2007; Lee et al., 2011; Sanchez-Mut et al., 2013; Xin et al., 2010).

In order to consider the spatial regulation of cerebral cortex and to discern the DNA methylation profiles of grey and white matter, we focused on dorsolateral prefrontal cortex Brodmann area 9 (BA9) because it is involved in higher cognitive skills such as memory and cognition (Wood and Grafman, 2003), and because it is also frequently altered in neurological diseases, including Alzheimer’s disease (AD) (Dournaud et al., 1995; Grousselle et al., 1998; Igarashi et al., 2011), Parkinson’s disease (PD) (Dumitriu et al., 2012; Lanoue et al., 2013; Lanoue et al., 2010), Huntington’s disease (HD) (Hoss et al., 2015; Labadorf et al., 2015; Mina et al., 2016), as well as mood and bipolar disorders (Brooks et al., 2010; Fitzgerald et al., 2006; Whitfield et al., 2015). We performed a whole genome bisulfite sequencing analysis (WGBS) of grey and white matter providing a comprehensive catalogue of DNA methylation profiles at base-pair resolution. In order to avoid inter-individual variability, we acutely dissected donor-matched grey and white matter from postmortem samples. Importantly, we validated the biological significance of these findings using an independent cohort of prefrontal cortex samples by Infinium 450K array and pyrosequencing.

Methods
Human samples – Post-mortem tissues were obtained from the Institute of Neuropathology Brain Bank (HUB-ICO-IDIBELL Biobank) following the practices and expertise of BrainNet Europe Bank (http://www.brainnet-europe.org/) "Network of Excellence" funded by the European Commission in the 6th Framework Program "Life Science" (LSHM-CT-2004-503039). All samples were obtained in agreement with ethical issues and legislations defined by the European Union and after the approval of the local Ethic Committee. DNA was extracted from grey and white matter of dorsolateral prefrontal cortex (Brodmann area 9). The EZ DNA methylation kit (Zymo Research) was used for bisulphite conversion of all DNA samples used as previously described (Bibikova et al., 2011).

Previously reported WGBS gray and white matter data from the same control donor were used (GSE47966 GSM1173772 – Grey matter – (Lister et al., 2013); GSM1279516 GSE52271–white matter – (Sanchez-Mut et al., 2016); female 64 years old). Five additional control grey and white matter were hybridized on 450K arrays (three females and two male aged 69.4 ±4.3 and four females and one male aged 64.4±9.7 respectively). A third set of ten control grey and ten white matter samples was used for pyrosequencing validation (seven females and three male aged 70.3 ±3.8 and eight females and two male aged 69.1±5.1 respectively).

Pyrosequencing – The set of primers for PCR amplification and sequencing (Supplementary table 1) were designed using the software PyroMark assay design version 2.0.01.15 (Qiagen); amplification primers hybridize with CpG-free sites to ensure methylation-independent reaction and one primer (opposite to the sequencing primer) is biotinylated to convert the PCR product to single-stranded DNA templates. We used 1ul of bisulphite-treated DNA for each PCR. We used (Bibikova et al., 2011)the Vacuum Prep Tool (Quiagen) to prepare single-stranded PCR products according to manufacturer’s instructions. Pyrosequencing reactions and methylation quantification were performed in a PyroMark Q24 System version 2.0.6 (Qiagen).

Whole-genome bisulfite sequencing data processing – Sequence alignment and methylation calling were performed with version 0.7.4 of Bismark software (Krueger and Andrews, 2011). We used hg19 as reference genome and retrieved genomic information from Biomart (Haider et al., 2009). SAM/BAM and BED files handling was done using SAMtools (Li et al., 2009), bedtools (Quinlan and Hall, 2010) and Tabix (Li, 2010). Statistical analysis and graphic representation was performed with R (http://www.R-project.org) and libraries multicore and ggplot2. We defined the promoter region as 2 kb flanking the transcription start site (TSS). TSS was considered to be the upstream-most base of all the transcripts of the gene. Smoothed methylation profiles were generated following the BSmooth pipeline (Hansen et al., 2012).

CpG distance correlation – We assessed distance correlation of CpG in close proximity gathering information about methylation and relative distance up to 2000 bases away from all CpG sites and correlating pair-wise methylation at single CpG sites for each relative distance.
Detection of differentially methylated regions – Differentially methylated regions (DMR) were identified as regions with more than 5 consecutive CpG sites that showed a consistent difference between the 95% CI of the smoothed methylation profile (Hansen et al., 2012).

Detection of histone marks – We obtained human histone marks for human adult brain inferior temporal lobe available at the ENCODE project (GSE17312) and CCAT peak caller (Xu et al., 2010) with the standard parameter for histone marks.

Microarray-based DNA methylation analysis with Infinium 450k array – The Infinium methylation assay was carried out as described previously (Bibikova et al., 2011). P-values for probes covering and following consistent tendencies with the identified DMRs were obtained using Kruskal–Wallis test and false discovery rate (FDR<0.05) adjustment for multiple comparisons.

RNA expression profiling – RNA expression profiling was performed following the Pico Profiling method (Gonzalez-Roca et al., 2010). Each sample analyzed by WGBS was also hybridized to a Human Gene ST 1.0 Arrays (Affymetrix). The arrays were scanned in a GeneChip Scanner 3000 (Affymetrix). CEL files were generated from DAT files using GCOS software (Affymetrix). To generate the log2 expression estimates, overall array intensity was normalized between arrays and the probe intensity of all probes in a probeset summarized to a single value using RMA (Robust Multichip Average) algorithm (Irizarry et al., 2003).

Disease-associated genes – AD-associated genes were obtained from AlzGene database (Bertram et al., 2007), PD-associated genes from PDGene (Lill et al., 2012), Amyotrophic lateral sclerosis-associated genes from ALSGene (Lill et al., 2011), and Multiple Sclerosis- associated genes from MSGene (Lill CM, 2012). Relative gene amounts was assayed using two-tailed Fisher’s exact test and the level of significance (alpha) set at p<0.05.

Results

WGBS patterns

We performed WGBS of grey and white matter from BA9 obtained from the same patient in order to avoid inter-individual variability. 19x10^7 paired-end reads corresponding to a global coverage of ~10x sequencing depth and interrogating ~94% of genome were successfully mapped and obtained for both samples. Closer inspection indicated that more than 10^9 single Cs (98%) were covered in both samples with 3.8x10^8 (6,2%) and 3.2x10^8 (5,2%) of reads methylated in grey and white matter respectively. In detail, 27.4x10^7 (79,3%) of CGs, 8.6x10^7 (3,8%) of CAs, 2.8x10^7 (1,3%) of CTs and 0.7x10^7 (0,5%) of CCs reads were methylated in grey matter whereas 28.5x10^7 of CGs (78,9%), 3.5x10^7 (1,5%) of CAs, 1.1x10^7 (0,5%) of CTs and 0.6x10^7 (0,3%) of CCs were methylated in white matter. Thus, we observed higher levels of non-CG methylation in grey matter (30,8% vs. 15,3% of total C methylated reads were in non-CG context) while CG methylation was similar in both samples (79,3% vs. 78,9%). Noticeably,
similar results were obtained in immunolabeled neuronal and glial cells suggesting that this pattern represents a general tendency in the brain (Lister et al., 2013).

Next, we asked how methylation values were distributed. As expected, we observed a similar bimodal distribution of CG methylation in both samples with ~88% of Cs methylated (>66% average methylation value) and ~5% of Cs unmethylated (<33% average methylation value). On the contrary, non-CG methylation showed a single distribution with most of Cs being unmethylated (~90% of total Cs with average methylation value lower than 33%) even in grey matter where overall methylation levels were higher (Supplementary figure 1).

In keeping with previous results (Heyn et al., 2012), we observed a significant correlation between neighboring CG methylation site levels (Supplementary figure 2). On the contrary, no correlation was observed for non-CG methylation. This might be explained by the fact that CG methylation shows a bimodal distribution and tends to form clusters whereas non-CG methylation tends to be more scattered (Hebestreit et al., 2013; Lister et al., 2013).

Next, we assessed the relation between CG and non-CG methylation. To solve the problem of the scattered pattern of non-CG methylation, we defined flanking windows of 500bp and 2kb around the CpG islands and interrogated the non-CG methylation in the neighborhoods of CG methylated and unmethylated islands. Interestingly, CG and non-CG methylation were correlated in grey and white matter samples with the former showing higher correlation values (Supplementary Figure 3; R= 0.51 p-val<0.001 and R = 0.21 p-val<0.001; Pearson correlation values for 500bp and 2kb windows respectively). In keeping with previous results (Guo et al., 2014; Kulis et al., 2015; Sun et al., 2016; Ziller et al., 2011), we observed a correlation between CG and non-CG methylation which was stronger in grey matter where neurons, with higher levels of non-CG methylation, are more abundant (Guo et al., 2014; Lister et al., 2013; Sun et al., 2016).

Considering all of the aforementioned, we then decided to focus all subsequent analyses on CG methylation for the following reasons: 1) The functional relevance of CG methylation in gene regulation has been unequivocally demonstrated (Vinson and Chatterjee, 2012); 2) the correlation between neighboring CG methylation site levels allows the use of smoothing algorithms increasing the power, accuracy and sensitivity to detect differentially methylated regions (DMRs) (Hansen et al., 2012); 3) high-density array platforms are designed preferentially for interrogating CG methylation sites with the non-CG methylation probes being almost negligible; and 4) in contrast to the scattered pattern of non-CG methylation, differentially methylated CG loci can be precisely determined facilitating their validation using locus-specific techniques such as pyrosequencing.

**Differentially methylated regions**

To assess CG methylation in more detail, we used previously published smoothing methods (Hansen et al., 2012) and found 26914 DMRs between grey and white matter of BA9 (mean
size ~ $10^2$-$10^3$ bp; Figure 1A). These DMRs tend to lie on functional relevant regions as expected by the regulatory role of CG methylation (3.4 OR; Bootstrap pval<0.001). 6447 DMRs were located on promoters (4.1 OR; Bootstrap pval<0.001), 18900 on gene bodies (1.6 OR; Bootstrap pval<0.001), 6292 on enhancers (7.4 OR; Bootstrap pval<0.001), 1046 on TFBS (7.1 OR; Bootstrap pval<0.001) and 4689 on regions without regulatory function (0.2 OR; Bootstrap pval<0.001) (Figure 1B). Similar enrichment of differentially methylation regions in enhancer regions have been previously observed (Kozlenkov et al., 2014; Sun et al., 2016), supporting the accuracy of our results.

**Epigenetic crosstalk between DNA methylation and histone marks**

After that, we interrogated the relation between DNA methylation and histone marks using the comprehensive epigenetic profiling of human brain samples within the Roadmap project (Bernstein et al., 2010) because there is an important crosstalk between both epigenetic marks (Cedar and Bergman, 2009), in particular also for higher cognitive functions (Miller and Sweatt, 2006). As expected, H3K4me3 and H3K9ac, which are enriched on active promoters, tend to be non-methylated (Supplementary figure 4). In contrast, H3K9me3, a marker of heterochromatin, shows higher levels of DNA methylation. Furthermore, H3K36me3, which marks gene bodies of active genes, shows high levels of DNA methylation and H3K27ac, H3K4me1 and H3K27me3, which mark enhancers and polycomb genes, respectively, show intermediate DNA methylation levels. Noticeably, most of the DMR co-localized with euchromatin marks (H3K4me1, H3K4me3, H3K36me3, H3K27ac; 8.3 OR; Bootstrap pval<0.001) suggesting that the majority of DNA methylation differences between grey and white matter affect regulatory active regions whereas heterochromatin regions remain largely unaltered (H3K9me3; 0.3 OR.; Bootstrap pval<0.001). DMRs tend to lie on euchromatin regions 27.6 times more than on heterochromatin regions (Bootstrap pval<0.001). In detail, 8307 DMRs were located on H3K4me3 (12.7 OR; Bootstrap pval<0.001), 11959 on H3K27ac (7.9 OR; Bootstrap pval<0.001), 1739 on H3K27me3 (7.1 OR; Bootstrap pval<0.001), 7836 on H3K4me1 (5.8 OR; Bootstrap pval<0.001) and 2185 on H3K36me3 (2.6 OR; Bootstrap pval<0.001; Figure 1B). Supporting these results, we observed an important overlap between DNA sensitive sites in human brain samples (ENCODE/Duke University; Frontal cerebrum) and DMRs (23.1 OR; Bootstrap pval<0.001), which suggests that DMR tend to be located on open chromatin regions. Closer inspection revealed that 1045 DMRs were located on bivalent domains (H3K4me3+H3K27me3) (17.6 OR; Bootstrap pval<0.001) indicating that bivalent domains were the most overrepresented histone mark category of the genome.

**Correlation between DNA methylation and RNA expression**

In addition, we examined the correlation between DNA methylation and RNA values using GeneChip human gene 1.0ST array data (Supplementary table 2). In keeping with the idea that promoter DNA methylation represses RNA expression (Kelly et al., 2012), promoter DNA methylation was negatively correlated with gene expression (Figure 1C; Grey matter R = -0.49;
White matter $R = -0.47$; Pearson correlations, p-val <0.001), supporting the biological significance of the observed DMRs. However, in contrast with previous observations (Zemach et al., 2010), no significant differences in RNA expression were observed in relation to gene body DNA methylation (data not shown).

**Validation of WGBS using 450K Illumina Platform and pyrosequencing**

In order to validate WGBS results, we took advantage of the Infinium Human Methylation 450 BeadChip Kit which interrogates more than 485,000 methylation sites per sample at single-nucleotide resolution. DNA methylation profiles obtained from both platforms were highly correlated ($R^2 = 0.957$ and $R^2 = 0.962$ for grey and white matter respectively). In detail, from 11860 DMRs identified by WGBS (Supplementary table 3), 4496 (38%) were also covered by 450K arrays and 3820 (85%) of those were technically validated between both technologies (Supplementary figure 5). For further biological validation and for avoiding subject-specific differences, we used an additional five grey and five white matter samples. Importantly, 4010 of covered DMRs (89.2%) showed consistent tendencies between independent samples (Supplementary figure 5) and 2806 (70 %) significant differences (Supplementary table ; FDR<0.05) reinforcing the biological importance of these findings. As expected, a clustering analysis of 450K CpGs covering DMRs showed a clear difference of DNA methylation between grey and white matter profiles in BA9 (Figure 2). Lastly, these data were also technically and biologically validated by pyrosequencing using a third set of ten grey and ten white matter samples. WGBS DMRs showing significant differences in the 450K arrays were sorted according to their CpG mean differences and the top ranked genes were interrogated by pyrosequencing. These results revealed similar DNA methylation differences between grey and white matter in WGBS, 450K data, pyrosequencing and RNA expression (Table 1, supplementary table 4, and supplementary figure 6).

**Enrichment of DMR on glial and neuronal genes**

Noticeably, we observed an important overlap of DMRs with promoters of neuronal (e.g., ENO2, TUBB3, MAP2 and MAPT) and glial genes (e.g., GFAP, GLUL, S100B and NOTCH1 for astrocytes and MBP, SOX10, MAG and MOG for oligodendrocytes: As expected, and in line with the negative correlation between promoter DNA methylation and gene expression, genes preferentially expressed in neurons were hypomethylated in grey matter whereas genes preferentially expressed in glial cells markers were hypomethylated in white matter (Figure 3A). Consistently, an outstanding number of genes differentially expressed genes between neurons, astrocytes, and oligodendrocytes (Cahoy et al., 2008) also shows differences in DNA methylation (Supplementary table 5).

Next, we compared our data with previously published and NCBI GEO accessible datasets investigating DNA methylation differences between neuron and glial cells. Our DMR list comprises more than 90% of genes covered by both platforms and differentially methylated in the 450K dataset GSE41826 consisting of 29 neuron and glia paired prefrontal cortex samples
(average DNA methylation between NeuN+ and NeuN− >5%; FDR<0.05) (Guintivano et al., 2013). Besides, since human and mouse DNA methylation differences between neurons and glial cells are highly conserved (Kessler et al., 2016), we compared our DMR list with the WGBS dataset GSM11737786-91 consisting of three paired cortex murine samples (Lister et al., 2013) reanalyzed by Kessler et al (Kessler et al., 2016). Our DMR list includes more than 70% of genes differentially methylated in Kessler et al. data (Supplementary table 1 Kessler et al., 2016), supporting the reliability and functional relevance of our observations.

In addition, we explored the functional particularities of grey and white matter genes with DMRs in promoters using GeneCodis tool (Carmona-Saez et al., 2007; Nogales-Cadenas et al., 2009; Tabas-Madrid et al., 2012). Differentially methylated genes were enriched in Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa and Goto, 2000) terms associated with neuron- and glia-specific processes such as regulation of actin cytoskeleton and axon guidance, Gap junctions, and Wnt signaling (Supplementary table 6). Similarly, Protein ANalysis THrough Evolutionary Relationships (PANTHER) (Thomas et al., 2003) terms such as synaptic vesicle trafficking, acetylcholine, oxytocine, GABA, and glutamate signaling genes were also enriched in DMRs between grey and white matter (Supplementary table 7). In this regard, these data are in line with the expected gene enrichments, presumably a direct consequence of the different composition of cells between grey (neuron-like) and white matter (glial-like).

Interestingly, KEGG and PANTHER analysis also identified a significant enrichment of DMR in genes associated with AD, PD, and HD as well as Amyotrophic lateral sclerosis (ALS) (Supplementary tables 6 and 7) indicating that an important fraction of differentially methylated genes between grey and white matter are associated with neurodegenerative diseases.

### Disease-associated genes differentially DNA methylated between grey and white matter

To corroborate these observations, we next took advantage of the AlzGene (Bertram et al., 2007), PDGene(Lill et al., 2012), ALSGene(Lill et al., 2011), and MSGene(Lill CM, 2012) datasets, which constitute independent repositories of genes associated with AD, PD, ALS, and Multiple Sclerosis (MS), respectively. From the identified 6447 DMRs on promoters, 190 corresponded to previously AD-associated genes (5.97 O.R.; Fisher-test; pval<0.05), 314 to PD-associated genes (3.81 O.R.; Fisher-test; pval<0.05), 142 to Multiple Sclerosis-associated genes (7.38 O.R.; Fisher-test; pval<0.05), and 35 to Amyotrophic lateral sclerosis-associated genes (28.93 O.R.; Fisher-test; pval<0.05) (supplementary table 8). A closer inspection revealed interesting DNA methylation profiles regarding these genes. For instance, several important genes associated with AD, such as Clusterin (CLU) (Harold et al., 2009) and microtubule-associated protein tau (MAPT) (Grundke-Iqbal et al., 1986), were hypomethylated in grey matter (Figure 3B). Similarly, genes associated with PD, such as Leucine-rich repeat kinase 2 (LRRK2) (Paisan-Ruiz et al., 2004) and PTEN Induced Putative Kinase 1 (PINK1) (Valente et al., 2004), were also hypomethylated in grey matter as well as Interleukin 1 Receptor Type 1 (IL1R1) (Schrijver et al., 1999) as well as TNF receptor superfamily member
1A (TNFRSF1A) (Cereda et al., 2008), which are associated with MS and ALS respectively (Figure 3B).

**Discussion**

In this study, we used a novel approach to determine the specific characteristics of grey and white matter divisions in BA9. Acutely dissected grey and white matter from BA9 was isolated and its DNA extracted. WGBS and Human Infinium Human Methylation 450 BeadChip Kit arrays were used to determine their DNA methylation profiles and the resulting data was further validated by pyrosequencing.

Despite that WGBS can provide data on CG and non-CG methylation, we decided to focus on CG methylation for several practical regions. First, the effect of CG methylation on gene regulation has largely been demonstrated (Vinson and Chatterjee, 2012). Second, CG methylation of neighboring CG sites is strongly correlated facilitating the application of smoothing algorithms that increase the power and significance of the results (Hansen et al., 2012). Third, high-density array platforms interrogate mostly CG methylation facilitating the cross-validation between platforms (Sandoval et al., 2011). And finally, in contrast to the scattered pattern of non-CG methylation, loci of differentially CG methylation can be precisely determined facilitating the validation of the data using alternative techniques such as pyrosequencing.

Doing so, we found an important enrichment of DMRs in functionally relevant genome regions, especially in enhancers, which are, in turn, strongly associated with cell-type specificity (Creyghton et al., 2010; Zhu et al., 2013), consistent with previous reports (Kozlenkov et al., 2014; Sun et al., 2016). Similarly, our data show a significant negative correlation between promoter DNA methylation and gene expression, consistent with the idea that promoter DNA methylation represses RNA expression (Kelly et al., 2012). Finally, our results were also in agreement with the expected differences in cell type markers in grey and white matter as well as between grey and white matter-specific molecular pathways, supporting the biological significance of the observed DMRs.

To our knowledge, our study is the first to provide a complete and detailed landscape of DNA methylation profiles between the anatomically distinct grey and white matter. Previous studies interrogating DNA methylation profiles of neuron and glial cells have used isolated NeuN-positive cells and ChIP-on-chip assays (Iwamoto et al., 2011; Labonte et al., 2013). ChIP-on-chip assays consist of immunoprecipitating DNA using antibodies that recognize 5mC or MBD proteins, a technique that is known to introduce a certain bias of sequence recognition and to be limited in resolution imposed by the array. Therefore, this technique can only represent a partial view of *bona fide* DNA methylation profiles. Others have used restriction enzyme-based methods (Kozlenkov et al., 2016; Montano et al., 2013; Sun et al., 2016) or Infinium HumanMethylation450 BeadChip arrays (Guintivano et al., 2013; Kozlenkov et al., 2014; Kozlenkov et al., 2016) that provide a broader representation of the genome but are biased to
CpG-rich regions. Circumventing this problem, Lister and collaborators have recently reported a WGBS study on isolated NeuN cells (Lister et al., 2013), which is not biased by immunoprecipitation and which analyzes virtually all nucleotides of the genome in a comparable resolution to our data. Nonetheless, all these studies are based on NeuN-stained isolation of cells, which does not take the spatial information or the anatomical specialization between different brain regions into consideration. Yet, neuron and glial cells show characteristic spatial expression profiles according to specific anatomical structures (Bohland et al., 2010; Ko et al., 2013) which are also reflected at DNA methylation profiles (Davies et al., 2012; Hernandez et al., 2011; Ladd-Acosta et al., 2007; Lee et al., 2011; Sanchez-Mut et al., 2013; Xin et al., 2010).

Notwithstanding, our approach might be limited by imperfections in dissection and the impurity of the grey and white matter divisions and the limited number of samples analyzed. Furthermore, standard bisulfite conversion cannot recognize hydroxymethylation (Huang et al., 2010), an emerging epigenetic modification in the brain with various functions including cognition (Massart et al., 2014; Mellen et al., 2012), nor can we capture the highly dynamic nature of DNA methylation, including DNA methylation changes associated with learning and memory (Graff and Mansuy, 2008; Lubin et al., 2011; Miller and Sweatt, 2006; Numata et al., 2012).

Interestingly, our analysis has revealed an outstanding enrichment of DMR in genes associated with several neurological and neurodegenerative diseases. In particular, we have observed an enrichment of DMRs in genes previously associated with AD, PD, HD, MS and ALS. AD- and PD-related genes such as CLU and MAPT, and LRRK2 and PINK1 are differentially methylated between grey and white matter, as well as IL1R1 and TNFRSF1A, which are associated with MS and ALS respectively. In line with the repressive effect of promoter DNA methylation, these genes tend to be hypermethylated (Figure 3B) and less expressed (Supplementary table 2) in white matter, where a higher proportion of glial cells are observed. In follow-up studies, it will be interesting to determine whether these genes are differentially methylated in brain regions particularly affected in neurodegenerative diseases. First attempts along this line, exemplified by MAPT DNA methylation, have shown a complex scenario in which the specific brain region DNA methylation patterns can be influenced by the genetic background and the disease status (Coupland et al., 2014).

These observations indicate that genes associated with neurodegenerative diseases are differentially regulated between grey and white matter, which are in turn differentially affected in such diseases, reinforcing the necessity of considering the cellular and spatial compartmentalization in neurodegenerative studies. In addition, these data also suggest that such genes play a major role in grey matter and, by extension, in neurons, while other genes showing the opposite relation might need to be interpreted as glia-relevant genes. Also, since neurodegenerative diseases are primarily affecting particular cell types and specific brain regions, an alteration of the cellular ratios and spatial distribution of their specific epigenetic patterns are expected. As a consequence, a direct comparison between control and disease-
affected tissues should not only result in alterations underlying neurodegeneration diseases but also in differences between cell patterns, that can be present even in studies based on neuron/glia fluorescence-activated cell sorting (FACS), where a minimal contamination of cells cannot be completely discarded. Therefore, previous studies investigating changes in DNA methylation in neurodegeneration in genes showing differences between grey and white matter, or neuron and glial cells, ought to be carefully interpreted in light of our observations presented here (for a detailed revision of genes previously associated with AD see (Sanchez-Mut and Graff, 2015), PD see (Feng et al., 2015), MS see (Li et al., 2016), and ALS see(Paez-Colasante et al., 2015)).

Taken together, the results presented here offer a complete and detailed landscape of DNA methylation profiles of grey and white matter in BA9 that constitutes an important resource to assist future studies determining cell type-specific (epigenetic) differences pertinent (or not) to this region of the cortex. Additionally, our data may also be relevant to unmask and better understand epigenetic alterations underlying neurological and neurodegenerative diseases, and to assist case-control studies where cell type compositions are not considered.
Acknowledgments

This work was supported by the European Research Council (ERC-2015-StG 678832), by the Swiss National Science Foundation, the National Competence Center for Research SYNAPSYS, the SYNAPSIS Foundation, the Béatrice Ederer-Weber Stiftung, the Fosshied Foundation, and the Alzheimer’s Association (NIRG-15-363964). HH is a Miguel Servet (CP14/00229) researcher funded by the Spanish Institute of Health Carlos III (ISCIII). ME is an ICREA Research Professor. JG is an MQ fellow and a NARSAD Independent Investigator.

Author contributions

JVSM, HH, IF, ME, and JG conceived the idea of this project. JVSM, and HH performed the experiments. EV, SM, SS, IF, JS, and RDM contributed to the analysis of the data. JVSM, ME, and JG wrote the manuscript with inputs from all the other authors.
References


Hoss AG, Labadorf A, Latourelle JC, Kartha VK, Hadzi TC, Gusevja JF, MacDonald ME, Chen JF, Akbarian S, Weng Z, Vonsattel JP, Myers RH. 2015. miR-10b-5p expression in
Huntington's disease brain relates to age of onset and the extent of striatal involvement. BMC medical genomics 8:10.


This article is protected by copyright. All rights reserved.


Figure 1. WGBS of Grey (G145) and White matter (W145) DMRs. (A) Circos representation of DMR between Grey and White matter samples. Outer circle represents average levels for all of DMRs in 50 Mbp-wide windows. Inner circle represents average levels for all of DMRs in 200 Kbp-wide windows. Bars indicate the magnitude of the difference between the G145 and the W145 samples for each window (color scale). (B) DMR distribution among different genomic regions, left. DMR overlapping with different histone marks, right. Observed DMRs refers to total number of DMRs obtained. Expected DMRs refers to expected DMR distribution randomly permuting their locations among the genome. (C) DNA methylation and gene expression correlation. DNA promoter methylation and gene expression correlation for each of samples analyzed (Grey matter R = -0.49; White matter R = -0.47; Pearson correlations, p-val <0.001).

Figure 2. Heatmap clustering of 450K data between grey and white matter samples. CpGs covering WGBS DMRs were selected to cluster five grey and five white samples. Red and green colors indicate high and low levels of DNA methylation respectively.

Figure 3. DMR examples of relevant genes. (A) DMRs overlapping promoters of cell type markers. (B) DMRs overlapping promoters of disease-associated genes. Smoothing values for grey and white matter are represented by red and blue lines respectively. Dots represent single sequencing reads, and light red and blue areas represent the confidence interval of the data.
Figure 1. WGBS of Grey (G145) and White matter (W145) DMRs. (A) Circos representation of DMR between Grey and White matter samples. Outer circle represents average levels for all of DMRs in 50 Mbp-wide windows. Inner circle represents average levels for all of DMRs in 200 Kbp-wide windows. Bars indicate the magnitude of the difference between the G145 and the W145 samples for each window (color scale). (B) DMR distribution among different genomic regions, left. DMR overlapping with different histone marks, right.

Observed DMRs refers to total number of DMRs obtained. Expected DMRs refers to expected DMR distribution randomly permuting their locations among the genome. (C) DNA methylation and gene expression correlation. DNA promoter methylation and gene expression correlation for each of samples analyzed (Grey matter R = -0.49; White matter R = -0.47; Pearson correlations, p-val <0.001).
Figure 2. Heatmap clustering of 450K data between grey and white matter samples. CpGs covering WGBS DMRs were selected to cluster five grey and five white samples. Red and green colors indicate high and low levels of DNA methylation respectively.

257x190mm (300 x 300 DPI)
Figure 3. DMR examples of relevant genes. (A) DMRs overlapping promoters of cell type markers. (B) DMRs overlapping promoters of disease-associated genes. Smoothing values for grey and white matter are represented by red and blue lines respectively. Dots represent single sequencing reads, and light red and blue areas represent the confidence interval of the data.

258x190mm (300 x 300 DPI)
Table 1. Panther terms enrichments on genes differentially methylated between grey and white matter.

<table>
<thead>
<tr>
<th>Panther terms</th>
<th>Genes</th>
<th>Hyp*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wnt signaling pathway</td>
<td>136</td>
<td>4.41E-25</td>
</tr>
<tr>
<td>Endothelin signaling pathway</td>
<td>50</td>
<td>7.42E-18</td>
</tr>
<tr>
<td>Angiogenesis</td>
<td>79</td>
<td>8.82E-17</td>
</tr>
<tr>
<td>EGF receptor signaling pathway</td>
<td>64</td>
<td>4.41E-16</td>
</tr>
<tr>
<td>PDGF signaling pathway</td>
<td>68</td>
<td>1.25E-15</td>
</tr>
<tr>
<td>Integrin signaling pathway</td>
<td>78</td>
<td>1.30E-15</td>
</tr>
<tr>
<td>Heterotrimeric G-protein signaling pathway-Gq alpha and Go alpha mediated pathway</td>
<td>61</td>
<td>1.35E-15</td>
</tr>
<tr>
<td>Inflammation mediated by chemokine and cytokine signaling pathway</td>
<td>89</td>
<td>1.82E-14</td>
</tr>
<tr>
<td>Heterotrimeric G-protein signaling pathway-Gi alpha and Gs alpha mediated pathway</td>
<td>71</td>
<td>4.92E-13</td>
</tr>
<tr>
<td>FGF signaling pathway</td>
<td>57</td>
<td>6.68E-13</td>
</tr>
<tr>
<td>Apoptosis signaling pathway</td>
<td>56</td>
<td>1.29E-12</td>
</tr>
<tr>
<td>Alzheimer disease-amyloid secretase pathway</td>
<td>41</td>
<td>9.12E-13</td>
</tr>
<tr>
<td>p53 pathway</td>
<td>45</td>
<td>2.19E-12</td>
</tr>
<tr>
<td>Muscarinic acetylcholine receptor 1 and 3 signaling pathway</td>
<td>36</td>
<td>1.14E-11</td>
</tr>
<tr>
<td>Oxytocin receptor mediated signaling pathway</td>
<td>35</td>
<td>1.59E-11</td>
</tr>
<tr>
<td>Ras Pathway</td>
<td>40</td>
<td>3.06E-11</td>
</tr>
<tr>
<td>Histamine H1 receptor mediated signaling pathway</td>
<td>29</td>
<td>5.28E-11</td>
</tr>
<tr>
<td>Thyrotropin-releasing hormone receptor signaling pathway</td>
<td>35</td>
<td>5.96E-11</td>
</tr>
<tr>
<td>Huntington disease</td>
<td>58</td>
<td>2.04E-10</td>
</tr>
<tr>
<td>TGF-beta signaling pathway</td>
<td>46</td>
<td>4.04E-10</td>
</tr>
<tr>
<td>Cadherin signaling pathway</td>
<td>60</td>
<td>3.01E-09</td>
</tr>
<tr>
<td>VEGF signaling pathway</td>
<td>33</td>
<td>3.31E-09</td>
</tr>
<tr>
<td>Alzheimer disease-presenilin pathway</td>
<td>53</td>
<td>3.57E-09</td>
</tr>
<tr>
<td>5HT2 type receptor mediated signaling pathway</td>
<td>35</td>
<td>4.01E-09</td>
</tr>
<tr>
<td>Notch signaling pathway</td>
<td>24</td>
<td>7.24E-09</td>
</tr>
<tr>
<td>GABA-B receptor II signaling</td>
<td>24</td>
<td>1.61E-08</td>
</tr>
<tr>
<td>Metabotropic glutamate receptor group III pathway</td>
<td>33</td>
<td>7.70E-08</td>
</tr>
<tr>
<td>Beta2 adrenergic receptor signaling pathway</td>
<td>25</td>
<td>8.75E-08</td>
</tr>
<tr>
<td>Beta1 adrenergic receptor signaling pathway</td>
<td>25</td>
<td>8.75E-08</td>
</tr>
<tr>
<td>Parkinson disease</td>
<td>40</td>
<td>9.87E-08</td>
</tr>
<tr>
<td>PI3 kinase pathway</td>
<td>28</td>
<td>1.37E-07</td>
</tr>
<tr>
<td>Muscarinic acetylcholine receptor 2 and 4 signaling pathway</td>
<td>29</td>
<td>4.20E-07</td>
</tr>
<tr>
<td>Alpha adrenergic receptor signaling pathway</td>
<td>16</td>
<td>4.94E-07</td>
</tr>
<tr>
<td>Metabotropic glutamate receptor group II pathway</td>
<td>25</td>
<td>4.95E-07</td>
</tr>
<tr>
<td>B cell activation</td>
<td>29</td>
<td>1.64E-06</td>
</tr>
<tr>
<td>Synaptic vesicle trafficking</td>
<td>17</td>
<td>2.01E-06</td>
</tr>
<tr>
<td>Metabotropic glutamate receptor group I pathway</td>
<td>16</td>
<td>2.88E-06</td>
</tr>
<tr>
<td>Oxidative stress response</td>
<td>25</td>
<td>5.88E-06</td>
</tr>
<tr>
<td>Ionotropic glutamate receptor pathway</td>
<td>23</td>
<td>5.90E-06</td>
</tr>
<tr>
<td>p53 pathway feedback loops 2</td>
<td>24</td>
<td>5.96E-06</td>
</tr>
<tr>
<td>Histamine H2 receptor mediated signaling pathway</td>
<td>15</td>
<td>8.78E-06</td>
</tr>
<tr>
<td>Interferon-gamma signaling pathway</td>
<td>17</td>
<td>1.55E-05</td>
</tr>
<tr>
<td>Nicotinic acetylcholine receptor signaling pathway</td>
<td>36</td>
<td>2.73E-05</td>
</tr>
<tr>
<td>Cytoskeletal regulation by Rho GTPase</td>
<td>30</td>
<td>2.91E-05</td>
</tr>
<tr>
<td>Axon guidance mediated by semaphorins</td>
<td>13</td>
<td>4.00E-05</td>
</tr>
<tr>
<td>T cell activation</td>
<td>32</td>
<td>4.26E-05</td>
</tr>
<tr>
<td>Circadian clock system</td>
<td>8</td>
<td>4.89E-05</td>
</tr>
<tr>
<td>Toll receptor signaling pathway</td>
<td>22</td>
<td>5.00E-05</td>
</tr>
<tr>
<td>Hypoxia response via HIF activation</td>
<td>15</td>
<td>6.28E-05</td>
</tr>
<tr>
<td>FAS signaling pathway</td>
<td>18</td>
<td>7.98E-05</td>
</tr>
</tbody>
</table>
Using whole genome bisulfite sequencing and high throughput arrays, the authors characterize the DNA methylation patterns of white and grey matter in the prefrontal cortex. Genes differentially methylated were found to be enriched in molecular pathways highly relevant for neuron/glia function and neurodegenerative diseases.
Using whole genome bisulfite sequencing and high-throughput arrays, the authors characterize the DNA methylation patterns of white and grey matter in the prefrontal cortex. Genes differentially methylated were found to be enriched in molecular pathways highly relevant for neuron/glia function and neurodegenerative diseases.