Specific Methyl-CpG Configurations Define Cell Identity through Gene Expression Regulation

Teresa Improda [1,†], Valentina Morgera [1,†] Maria Vitale [1], Lorenzo Chiariotti [2], Fabiana Passaro [2], Antonia Feola [1], Antonio Porcellini [1], Mariella Cuomo [2,*] and Antonio Pezone [1,*]

[1] Dipartimento di Biologia, Complesso Universitario di Monte Sant'Angelo, Università di Napoli "Federico II", Napoli 80126, Italy

[2] Dipartimento di Medicina Molecolare e Biotecnologie Mediche, Università degli Studi di Napoli "Federico II", Napoli 80131, Italy

[*] These authors equally contributed to this work.

[†] Correspondence: mariella.cuomo@unina.it (M.C.); antonio.pezone@unina.it and (A.P.)

Abstract: Cell identity is determined by chromatin structure and the profiles of gene expression, which are dependent on chromatin accessibility and DNA methylation of regions critical for gene expression, such as enhancers and promoters. These epigenetic modifications are required for mammalian development and are essential for the establishment and the maintenance of the cellular identity. DNA methylation was once thought to be a permanent repressive epigenetic mark, but systematic analysis in various genomic contexts reveals a more dynamic regulation than previously thought. In fact, both active DNA methylation and demethylation occur during cell fate commitment and terminal differentiation. To link methylation signatures of specific genes to their expression profiles, we have determined the methyl-CpG configurations of the promoters of five genes switched on and off during murine postnatal brain differentiation by bisulfite-targeted sequencing. We report here the structure of significant, dynamic and stable methyl-CpG profiles associated with silencing or activation of expression of genes during neural stem cell and brain postnatal differentiation. Strikingly, these methylation cores mark different mouse brain areas and cell types derived from the same areas during differentiation.

Keywords: methyl-CpG; DNA methylation; gene expression; cell identity

1. Introduction

DNA methylation, a covalent alteration of the cytosines in genomic DNA, primarily affects the CpG dinucleotide in vertebrates and is frequently linked to long-term transcriptional repression [1,2]. The DNA methyltransferases DNMT1 and DNMT3A/B/L maintain and establish methyl-CpG, respectively [2]. Conversely, methyl-CpG can be eliminated through dilution, passive demethylation by inactivation of DNMTs, or active demethylation by the Ten-eleven translocation (TET) and base excision repair (BER) enzymes [3]. However, several recent studies suggest that the roles of methyl-CpG and its oxidized derivatives, including hydroxymethylation (hydroxy-methyl-CpG), are more complex than previously believed [4–7]. Methyl-CpG can act as a cellular epigenetic trait involved in embryogenesis, cellular differentiation, and reprogramming due to its ability to be passed down through cell divisions. The methyl-CpG profiles can now be examined genome-wide and at single base resolution thanks to the technological advancements. These whole-genome methodologies have been used in studies that have significantly modified our understanding of the genomic distribution and spatiotemporal dynamics of methyl-CpG in several experimental models [8,9]. Our understanding of how methyl-CpG profiles affect genome expression is rapidly developing thanks to the combination of high definition of methyl-CpG profiling and functional studies in various model systems.

In a previous study, we presented a new tool to analyze DNA methylation in complex DNA sequence populations, i. e. MethCoresProfile. This program identifies and tracks epiallele (alleles different only by methylation) families in complex cell populations such as those derived from the

brain or other organs during postnatal differentiation [10]. MethCoresProfiler sorts out populations of DNA sequences derived from specific loci harboring a nucleus or core of CpGs, which are stably methylated in a relevant fraction of sequences in the population. These methylated cores define and mark the epialleles derived from a single ancestor (which generates the epiallele family). With time, CpGs surrounding the CpGs nucleus or core, change the methylation status, because they are not subjected to selection and contribute to the polymorphism and heterogeneity of methylation profiles of all the epialleles present in the population. It is worth noting that in complex mixtures of methylated DNA molecules, the frequency of the individual epialleles is never statistically significant because they are formed by stable (core) and unstable CpGs [10]. In simpler words, the methylation cores define CpGs subjected to selection and characterize a stable core in families of epialleles. The same applies to negatively selected epialleles, which disappear with time in the population of sequences. In both cases, positive and negative selection of epiallele families implies selection on the expression of the gene(s) represented in each family. With the publicly available MethCoresProfiler, we were able to generate high-resolution and dynamic epigenetic maps of specific loci [10], for example, during neurodevelopment. In fact, during the development of the brain, stem cell fate decisions are tightly correlated with their epigenetic status and expression profiles [11]. To map and identify DNA methylation cores during the neurodevelopment, we performed targeting-bisulfite amplicon sequence analysis of five genes involved in murine neural stem cell differentiation (mESC) to identify and characterize methyl-CpG configurations associated with the specific timing of expression of these genes during mouse brain postnatal differentiation. Our data demonstrate that specific methyl-CpGs cores in the promoters of these genes are tightly associated with induction or activation of transcription during brain differentiation in both the brain areas and isolated cells. We focused our analysis to the promoters, excluding enhancers, because the location of the promoters, not of the enhancers, at the 5' end of the target gene is invariant.

2. Results

2.1. Specific Methylated Cores Mark the Promoters of Genes Involved in Murine Postnatal Brain Differentiation

Cellular differentiation is characterized by reduction of the proliferation potential and the expression of markers defining the cell type. The reduction of the developmental potential is driven by epigenetic changes that prevent the risks linked to the expression of non-lineage-related genes in adult cells [12-14]. In this context, the specific role of CpG methylation of the regulatory gene regions that control cell differentiation is still debated. Moreover, the methylated segments of differentiation or stemness genes are highly polymorphic and heterogeneous because, when transcribed, they display stochastic hydroxy-methyl-CpGs, which greatly dilute stable methylated and inheritable sequences subjected to selection [10]. Indeed, in several cases where methylation was measured at the single cell level, a remarkable degree of methylation polymorphism was found in all cell types with no evidence of clonal and stable epialleles [15,16]. The considerable degree of heterogeneity is also demonstrated by measurements of the entropy index, i.e., the number of individual species within a population of epialleles, according to Shannon [17]. Furthermore, limited sampling and variability in mean methylation of individual CpGs in the sequence populations introduce additional bias into the analysis. As mentioned above, we have solved this problem by extracting common methylation signatures or cores in the population of heterogeneous sequences. In this way, we can identify stable families of epialleles and study their evolution over time, for example, during the differentiation of stem cells [10].

We have applied this approach to study the epigenetic evolution of specific genes, induced or repressed, during murine neural stem cell differentiation (mESC). We analyzed the promoter methylation configuration(s) of genes markers of pluripotency (such as *NANOG*) [18], or of neuronal (*TUBB3*, *DDO*) [19,20], or glial cells (*GFAP*) [21], or of differentiation or production of brain-derived neurotrophic factor (*BDNF*) [22] at various time points during postnatal mouse brain differentiation. We report here the identification of stable gene-specific methylation signatures that characterize

epiallele families of several genes associated with silencing or induction of gene expression during mouse brain postnatal differentiation.

2.1.1. *NANOG* Repression Is Associated with *de novo* Formation of a Methylated Core in the Promoter

NANOG is a key regulator of self-renewal and maintenance of pluripotency in undifferentiated embryonic stem cells. Each embryonic stem cell (ESC) in a population displays fluctuating Nanog levels that affect cell fate specification. During stem cell differentiation, *NANOG* is progressively silenced. We examined the methylation status of the *NANOG* promoter (Figure 1A) during ESC differentiation (Figure 1B). We found that all the cytosines in the region underwent progressive methylation during differentiation with different rates (Figure 1C), generating heterogeneous and polymorphic methylation profiles of the promoter molecules (Figure 1D). Using the tool indicated above (MethCoresProfiler) we identified in this heterogenous population of promoter sequences a stable CpG core tightly associated with progressive *NANOG* silencing in differentiating cells. At T2-4, the **NANOG** promoter shows the first round of methylation at several CpGs, but the stable and statistically relevant methylation core appears and stabilizes at T8-14 (Figure 1E-F). In addition, we note a transient methylation at CpGs 80, which is rapidly substituted by the stable core of CpGs 365-375. With time the entire promoter region becomes methylated in differentiated cells. We hypothesize that CpGs 80 is OH methylated and with its demethylation at T8, favors the appearance of the core 365-375, which initiates the methylation of the promoter leading to the silencing of the gene (Figure 1E, F, G, H).
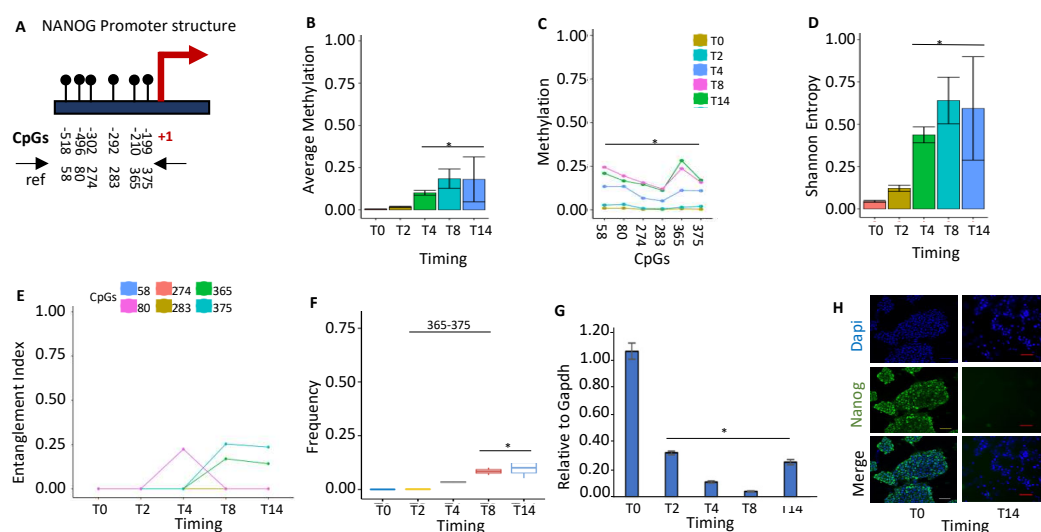
Figure 1. Methylation of *NANOG* promoter during stem cell differentiation. (A) Structure of the *NANOG* promoter. The black lines and circles on the diagram represent each CpG upstream of the transcription start site (TSS). (B) Average DNA methylation of the six CpGs shown in (A) at various time points (T0, T2, T4, T8, and T14). CpGs are shown as color-coded squares on the right. (C) Average CpG methylation is shown in (A). (D) The methylated molecules' Shannon entropy of the same samples and time intervals. (E) The methylated cores' composition and structure at various times during differentiation. A color code is used to identify each CpG on the right side of the panel. (F) Frequency of the methylated core in the whole population. (G) Gene expression analysis. (H) Immunofluorescence analysis of Nanog expression in undifferentiated ESCs (T0) and differentiated ESCs (T14). Scale bars: 50 μm. A pairwise comparison was performed with the student's t-test: *P < 0.05 versus T0.

2.1.2. *TUBB3* Promoter Methylation Does Nsot Control the Expression during Neuronal Differentiation

TUBB3 is almost exclusively expressed in neurons and is induced early during differentiation and reprogramming of neurons [23,24]. Examining the methylation status of the *TUBB3* promoter (Figure 2A), we identified a stable methylated nucleus at CpGs 22-35 (Figure 2E), which is present at the time 0 and does not significantly change over time (Figure 2B and 2C) during differentiation.

This segment upstream of the gene in mice and humans contains binding sites for several DNA-binding proteins, mainly located on the positive (+) strand [25]. This evidence suggests that *TUBB3* expression is not controlled by DNA methylation of the proximal promoter during neural stem cell differentiation (Figure 2G), as also confirmed by immunofluorescence (Figure 2H).
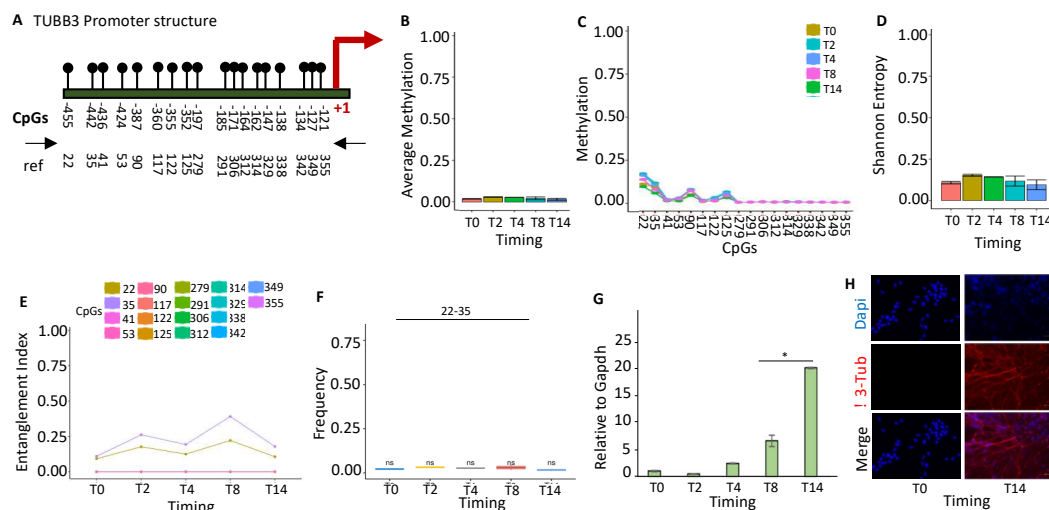
Figure 2. Methylation configuration of *TUBB3* promoter during stem cell differentiation. (A) Structure of the *TUBB3* promoter. The black lines and circles on the diagram represent each CpG upstream of the transcription start site (TSS). (B) Average methylation of the 18 CpGs shown in (A) at various stages of stem cell differentiation (T0, T2, T4, T8, and T14, color-coded squares on the right). Average CpG methylation in the population of molecules, in (C). (D) Shannon entropy of methylated molecules in the same samples and time intervals. (E) The methylated cores' composition and structure at various times during differentiation. A color code is used to identify each CpG on the right side of the panel. (F) Frequency of methylated core in the total population. (G) Gene expression analysis. (H) Immunofluorescence analysis of TUBB3 expression in undifferentiated (T0) and differentiated ESCs (T14). Scale bars: 50 μm. A pairwise comparison was performed with the student's t-test: *P < 0.05 versus T0.

2.1.3. *BDNF* Gene Activation Parallels the Loss of the Methylated Core at the Promoter during Brain Differentiation

Brain-derived neurotrophic factor (*BDNF*) is a member of the neurotrophin growth factor family that stimulates neurons' differentiation, maturation, and survival by suppressing apoptosis. *BDNF* also exerts a neuroprotective effect under adverse conditions [26]. We examined the epigenetic configuration of the *BDNF* promoter during differentiation and found a significant reduction of methylation (Figure 3A, Figure 3B). All CpGs in the promoter were found methylated in a small fraction of T0 cells, except CpGs 81-83, which constitute a common methylation core in the *BDNF* promoter present in the majority of undifferentiated cells (T0) (Figure 3). Induction of differentiation was associated with a sharp loss of methylation of all CpGs methylated at low frequency and of the CpGs 81-83 of the core at T0-T2 (Figure 3C). The loss of methylation reduced the epigenetic heterogeneity, and the Shannon entropy during differentiation (Figure 3D). However, our analysis reveals also that the methylated core, CpGs 81-83, present at T0 increased transiently at T2 and decreased at T8-T14 (Figure 3E,3F, 3G). The 81-83 CpGs were methylated at T0 and T2 in both strands

of the *BDNF* promoter, with a slight preference for the (-) strand (Figures S1-S2). The transient methylation rise of the 81-83 CpGs in the early phases of differentiation suggests a priming effect of this core on *BDNF* expression induced by differentiation, due to OH demethylation.
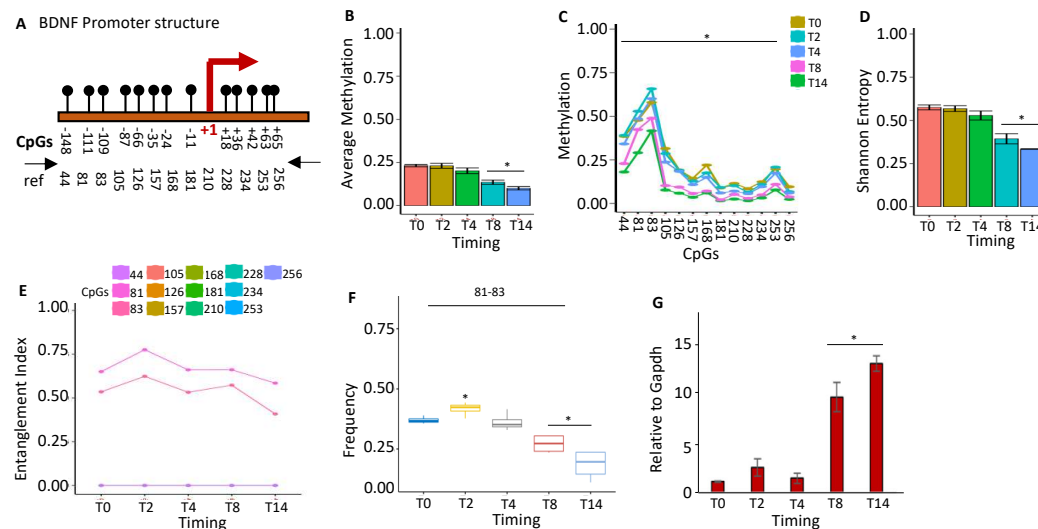
Figure 3. Methylation configuration of *BDNF* promoter during stem cell differentiation. (A) Structure of the *BDNF* promoter. The CpGs are represented by black circles and lines. (B) Average DNA methylation of the 13 CpGs shown in (A) at various stages of stem cell differentiation (T0, T2, T4, T8, and T14, color-coded squares on the right). (C) The average level of methylation for each CpG shown in A. (D) The methylated molecules' Shannon entropy for the same samples and time intervals. (E) The composition and structure of the methylated cores at various times. A color code is used to identify each CpG on the right side of the panel. (F) Frequency of methylated core in the total population. (G) Gene expression analysis. A pairwise comparison between each pair was performed with the student's t-test: *P < 0.05 versus T0.

2.1.4. *GFAP* Gene Activation Changes the Configuration of the Promoter Methylated Core

Glial fibrillary acidic protein (*GFAP*) is the protein in astrocytes of the primary intermediate filament (IF). In the human brain, *GFAP* isoforms display unique expression profiles, which suggest distinct functional roles. One isoform, GFAPδ, is expressed in proliferative radial glia precursors during human brain development. In humans, *GFAP* is a marker of neural stem cells [27]. We examined the methylation profile(s) of the *GFAP* promoter during mouse brain postnatal differentiation. We found that (Figure 4A), during ESC differentiation, all cytosines in the promoter sequence (Figure 4B) at low frequency undergo methylation (Figure 4C). In this context, we identified a stable methylation core at CpGs 35-84 in undifferentiated cells, which changed its structure at T4 (84-304) and stabilized at T8-T14 (Figure 4E). This nucleus (84-304 CpGs) becomes dominant in the whole population (Figure 4F) and marks the transcriptional *GFAP* activation (Figure 4G) [28], as also confirmed by immunofluorescence (Figure 4H).
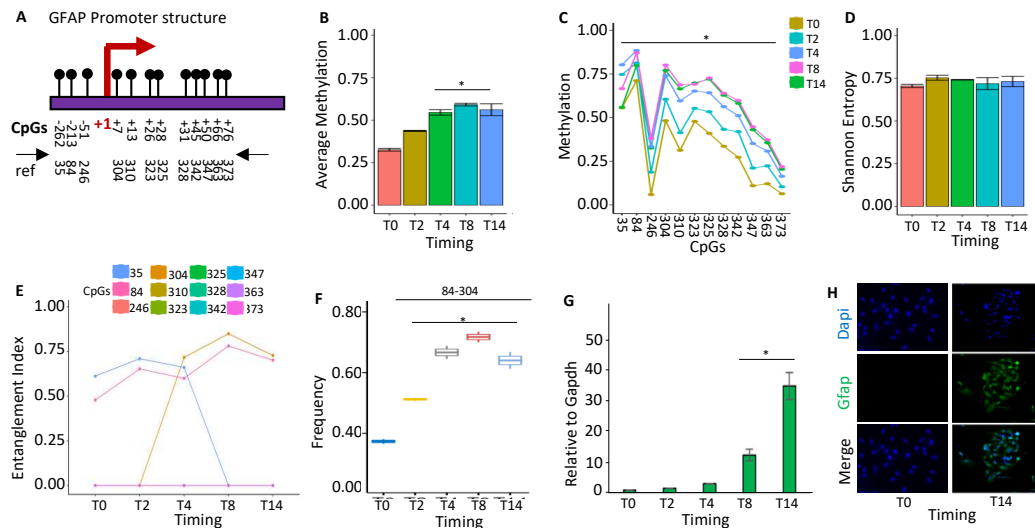
Figure 4. Methylation configuration of *GFAP* promoter during stem cell differentiation. (A) The *GFAP* promoter's structure. The CpGs are represented by black circles and lines. (B) Average methylation of the 12 CpGs depicted in (A) in cells at various time points during stem cell differentiation (T0, T2, T4, T8, and T14, color-coded squares on the right). (C) Average CpG methylation as displayed in (A). (D) The methylated molecules' Shannon entropy for the same samples and time intervals. (E) The composition and structure of the methylated cores at various times. A color code is used to identify each CpG on the right side of the panel. (F) Frequency of the methylated core in the total population. (G) Gene expression analysis. (H) Immunofluorescence analysis of GFAP expression in undifferentiated ESCs (T0) and differentiated ESCs (T14). Scale bars: 50 μm. A pairwise comparison was performed with the student's t-test: *P < 0.05 versus T0.

2.1.5. *DDO* Gene Promoter Methylation Profile and Expression during Postnatal Brain Differentiation

DDO (d-aspartate oxidase) levels are induced during brain differentiation and are inhibited in the adult brain, mirroring, reciprocally, d-aspartate levels, which are low in adults [29]. We examined the methylation status of the *DDO* promoter (Figure 5A), which shows a transient increase at T4 and returned to the baseline levels at T8-T14 (Figure 5B). This methylation involves all cytosines in the locus (Figure 5C) and correlates with a transient reduction of population heterogeneity at T4 (Figure 5D). We identified a methylation core in the *DDO* promoter at position 105-138-150 at T0-T2 which is fully methylated at T4 (105-138-150-226-293-343) and changes configuration at T8-T14 (105-138) (Figure 5E). This nucleus (105-138) significantly increases its frequency in the population up to T4 and then decreases (Figure 5F) marking the activation of the *DDO* transcription (Figure 5G) [30,31].
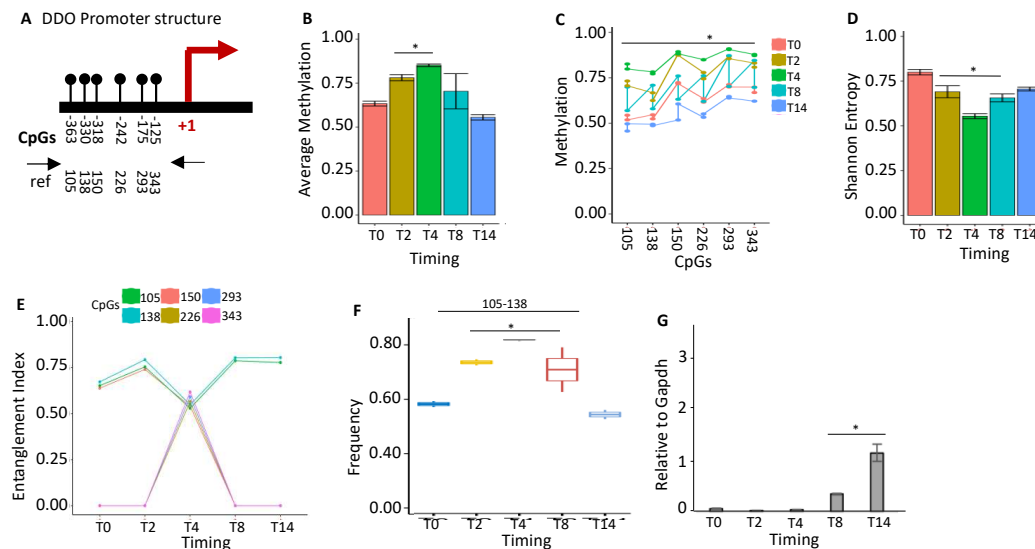
Figure 5. Methylation status of *DDO* promoters during stem cell differentiation. (A) The *DDO* promoter's structure. The CpGs are represented by black circles and lines. (B) The average methylation of the six CpGs is shown in (A) in the DNA extracted from stem cells at various time points (T0, T2, T4, T8, and T14, color-coded squares on the right). (C) The average level of methylation for each CpG found in the populations of molecules, as shown in (A). (D) The methylated molecules' Shannon entropy for the same samples and time intervals. (E) The composition and structure of the methylated cores at various times. A color code is used to identify each CpG on the right side of the panel. (F) Frequency of methylated core in the total population. (G) Gene expression analysis. A pairwise comparison was performed with the student's t-test: *P < 0.05 versus T0.

2.2. CpG Methylation Cores Define Cell Identity

Taken together, these data suggest that the methylation cores in the promoter mark the genes that determine the cell identity. To strengthen this hypothesis, we further explored the methylation cores in the region surrounding the promoter of the *DDO* gene in the whole brain during development *(accession number PRJEB16320)* [30, 31]. In the 3Kb region surrounding the *DDO* locus, analyzed by dividing it into seven amplicons, we found changes in methylation and methylated core structures only upstream of the *DDO* promoter (-1000 to +1) (Figures S3-S6) [30]. Moreover, we found the same methylated core (105-138) in region R4, which overlaps with the analyzed region in Figure 5 (Figure S5). To demonstrate that this core is associated with cell identity, we analyzed the methylation status of the *DDO* located in the R4 region in the DNA extracted from brain areas, several cell types isolated from the same brain areas, immortalized A1 neuronal cells, c-myc immortalized neurons at different passages and ESCs upon induction of neural differentiation [30] and gut (Figure 6A) [see materials and methods, 30]. Strikingly, this core and its components mark all brain areas and the cells isolated from the same brain areas [10]. The correlation between the presence of the *DDO* methylation core and the brain areas and/or cells isolated from the same areas is shown in Figure 6B, in which each core component is identified by the CpGs (105-138-150-226-293 relative to the *DDO* transcription start site) and a color code that marks the specific cell type and the mouse brain area(s) analyzed at different times during postnatal differentiation. Specifically, this tree displays several branches that link cells or brain areas to the specific *DDO* core segments: 1. cells (neural differentiated ES/Oligodendrocytes/Astrocytes) and cerebellum/prefrontal/striatum area, (red, CpG 105-138); 2. cortex (black, CpG 105-138-150-226-293 and brown, CpG 226-293); 3. cortex/microglia/undifferentiated ES (green, CpG 138-226); 4. hippocampus (blue, 105-138-150).

In conclusion, the phylogenetic tree in Figure 6 demonstrates that the *DDO* promoter core differentially marks the cells and the specific brain areas during postnatal differentiation [32,33,34,35].
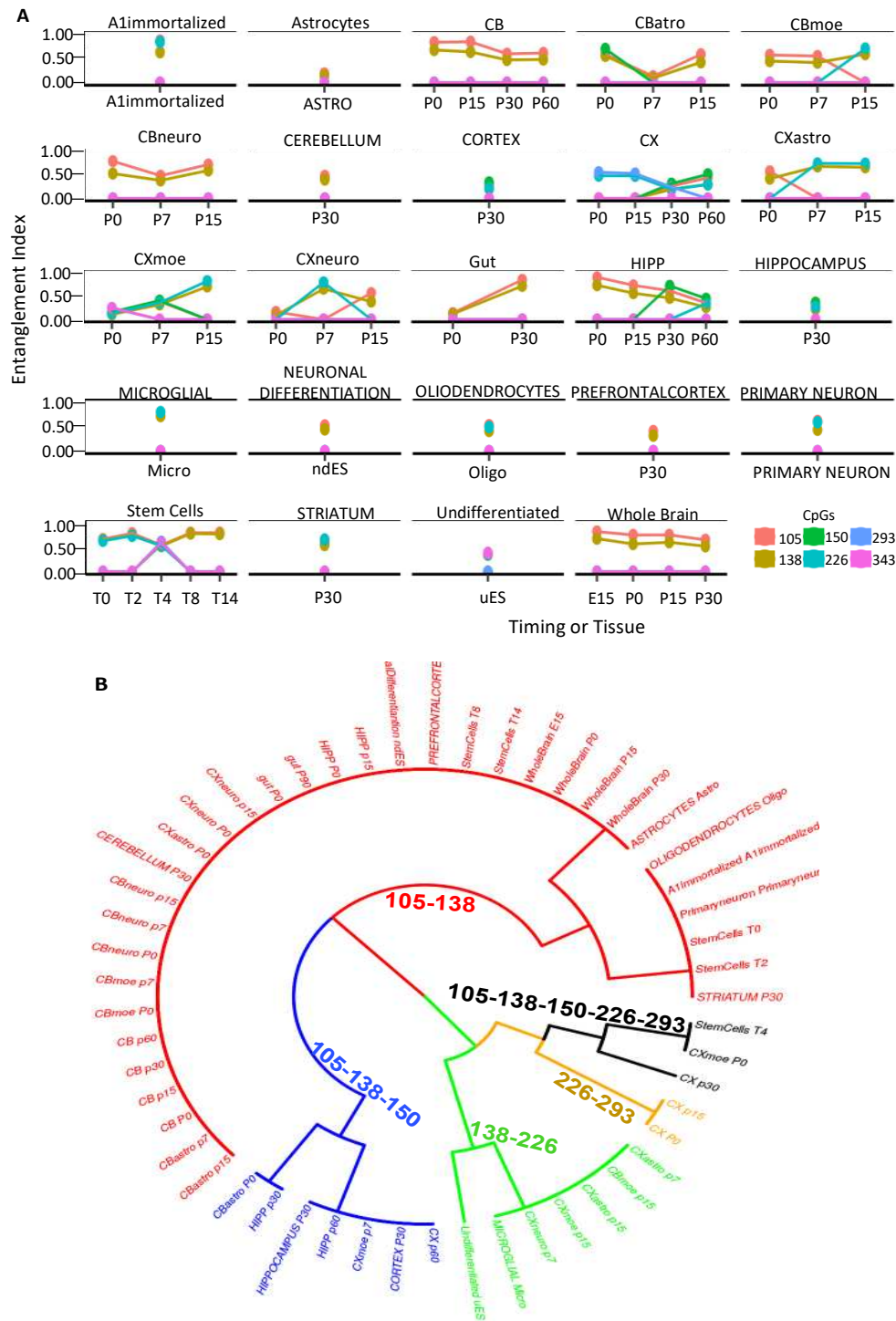
Figure 6. The methyl-CpG configurations of the *DDO* promoter define cell identity and brain areas. (A) *DDO* promoter methylated core structure and composition in fractionated cell populations and mouse brain regions at different times (Figure S5). (B) A phylogenetic tree describing the relationships between cell identities and the arrangement of methyl-CpGs. A color code is used to identify each CpG core on the right side of the panel.

3. Discussion

Specific memory mechanisms that regulate gene expression patterns epigenetically have evolved to establish and maintain cellular identity during development. Once determined, these lineage-specific expression profiles must be maintained across cell divisions, defining active or inactive gene expression states [32]. The most noticeable of these marks is the methylation of the

carbon-5 of cytosine (5mC), which has traditionally been thought incompatible with active transcription when located near or in gene regulatory regions. 5mC can modulate transcription factor binding [33] or induce the binding of specific 5mC-binding proteins, which can lead to the recruitment of co-repressor complexes to methylated target promoters.

The data we have shown demonstrate that heterogeneity of methylation profiles at the promoter sites of genes expressed or silenced during mouse postnatal brain differentiation is apparent, not substantial, because we identify stretches of 3 or 4 non-contiguous CpGs present in the majority of sequences analyzed. However, considering that in a promoter segment, at low levels every CpGs is transiently methylated and demethylated during transcription [4,8,10,15,16], every transcribed OH-methyl CpG is detectable as methyl-C when exposed to bisulfite [5]. By extracting significant non-contiguous CpGs methylated from pools of sequences, we identify methylated CpGs which are stable and highly frequent in the population (methylation cores or nuclei). These cores are statistically significant because they are present in molecules non-identical individually but derived from the same original precursor (epiallele family) [10]. We have identified several epiallele families in the promoters of 5 genes activated or repressed during postnatal mouse brain differentiation. These families are characterized by cores of 3- 4- 6- methylated CpGs that are found in various combinations in different cell types and brain areas. Progressive demethylation of these cores in the promoters of several genes involved in mouse brain postnatal differentiation is associated with the activation of transcription of specific cells and areas of differentiating mouse brain (*BDNF*, Figure 3). The contrary (progressive methylation), as in the *NANOG* promoter, is associated with silencing. However, we also find the presence of a promoter-methylated core (*GAP*) associated with the activation of transcription [21]. The best example of how a methylation core might associate with induction or timely inhibition of transcription is provided by the *DDO* promoter across 3Kb of a genomic region. *DDO* promoter displays a methylation core undergoing transient methylation early during differentiation (T4) followed by demethylation at late stages (T8-T14).

Moreover, we show that differentiated stem cells and neurons expressing *DDO* acquire variants from the same original methylated core (Figure 6A), which appears as identity marker of the precursor cells (Figure 6B). It is worth noting that the analysis with the current methods (single methylated CpG, entropy, or average methylation) do not reveal any stable methylation core (Figure S7).

We wish to note that our analysis identifies the same neuronal signature in the gut cells (Figure 6). This can be explained by the notion that the enteric nervous system (ENS) derives from the migratory capacity of the neural crest [34,35].

We are aware that the main limitation of this study is the circumstantial nature of the evidence associated with the cell type identity and methylation cores in neurodevelopment. Although we did not demonstrate that methylation promoter profiles *de facto* silence or activate transcription, the promoter methylation cores we have identified are tightly associated with the dynamic expression profiles of 4 genes important in developing and differentiating ES and mouse brain.

4. Materials and Methods

ESC Culture and Differentiation

E14Tg2a (BayGenomics) mouse ESCs were maintained as described elsewhere [36-38]. To induce neural differentiation, ESCs were plated onto gelatin-coated dishes at low density (1×103-5×103 cells/cm2) in the following differentiation medium: knockout DMEM sup-plemented with 10% KSR, 2 mM glutamine, 100 U/ml penicillin/streptomycin (Invitro-gen), 0.1 mM β-mercaptoethanol (Sigma).

RT-PCR

Total RNAs were extracted using Trizol (Invitrogen). RNA (2 μg/reaction) was reverse-transcribed using M-MuLV reverse transcriptase (Biolabs). The list of primers is reported in Supplementary Material Table S1.

Immunofluorescence

For immunofluorescence analysis, ESCs were fixed, permeabilized, and incubated with primary antibodies and an appropriate secondary antibody as previously described [39]. Nuclei were counterstained with DAPI (1:5000; Calbiochem). The following primary antibodies were used: anti-βIII Tubulin (1:400; Sigma); anti-GFAP (1:300; Sigma); an-ti-Nanog (1:400, Abcam). Alexa Fluor 594 or 488 secondary antibodies were used (1:400; Thermo Fisher Scientific). Cells were visualized using an inverted microscope (Leica Microsystems), and the images were captured with a digital camera (DFC365 FX; Leica Microsystems) using LAS-AF software (Leica Microsystems). Confocal images were ac-quired with LSM510META microscope (Carl Zeiss GmbH) using LSM510 software (Zeiss). After acquisition, the images were color corrected using the brightness, contrast, and color-balance commands applied to every pixel in each image.

DNA extraction

In accordance with the manufacturer's recommendations, DNA was prepared using the DNeasy® Blood & Tissue Kit from Qiagen in Hilden, Germany. Thermo Scientific's NanoDrop 2000 was used to check the DNA's quality. The 260/280 absorbance ratio was used to determine its quantity (Invitrogen, Q32850).

Bisulfite Treatment and Amplicon Library Preparation

EZ DNA Methylation Kit was used to perform the genomic DNA bisulfite treatment (Zymo Research) according to manufacturer's instructions. An amplicon library was sequenced by an Illumina Miseq Sequencer to determine the DNA methylation levels. The bisulfite conversion rate was estimated to be 98-99 percent. Table S1 in the Supplementary Material contains a list of primers.

Dataset Description of Figure 6

The data are available on the ENA database (accession number: PRJEB16320) [30].

All animals were derived from the Jackson Laboratory-provided C57BL/6J mice. All animal research was carried out as described in [30]. Whole brains were extracted from mice at various developmental stages, including E15, P0, P7, P14, P21, P30, and P60. From two mice, five brain areas were dissected (prefrontal cortex, cortex, hippocampus, cerebellum, and striatum) (P30).

Cortical neurons were isolated from the brains of C57BL/6J mouse embryos that were 17 days old, as described in [30]. Triturated tissues were plated in a culture medium, and ara-C (10 mM) was added within 48 hours of plating to prevent the growth of non-neuronal cells.

Mixed glial cells, purified microglia, oligodendrocytes, and astrocytes were prepared from the primary mouse as described in [30], obtaining IB4- FITC or OX42-positive cells, NG2-positive cells and GFAP-positive cells.

Differentiation of embryonic stem cells toward neurons and glia cells was done as previously described [36].

A1 mes c-myc (A1) is a cell line immortalized by infecting a primary mouse mesencephalon derived cells culture with a c-myc-carrying retroviral vector prepared from 11-day-old embryos (E11) as described in [30].

C57BL/6J mice were also used to obtain gut tissues from three newborn mice (P0 status) and three adult mice (P90 status) [38].

Sequence Handling

Paired-end reads in FASTQ format from the ENA database (accession number: PRJEB16320) and generated in our lab were merged using the PEAR (paired-end read merger) tool, with a minimum overlapping region of 40 nucleotides. Only reads with a mean quality score (Phred) greater than 33 and a length between 400 and 500 nucleotides were kept. The reads were then converted to FASTA format using the PRINSEQ (preprocessing and sequence information) tool. Reads in FASTA format were processed using ampliMethProfiler (available at

https://sourceforge.net/projects/amplimethprofiler/) with several quality filters to extract mCpG configurations in single DNA molecules. We only kept reads that were I 50% longer than the reference length, (ii) at least 80% sequence similarity of the primer with the corresponding gene, (iii) at least 98 percent bisulfite efficiency, and (iv) alignment of at least 60% of their bases with the reference sequences. All cytosines in the CpG sequence context have their methylation status coded as methylated (1) or unmethylated (2). Reads with ambiguous CpG dinucleotide calls (including gaps or A or G) were removed. The binary data was analyzed with the MethCoresProfiler [10] in stages.

Statistical Analysis

Average methylation data are expressed as means standard deviation. The unpaired Student t-test was used to make comparisons between two groups. Multiple comparisons were performed using 1-way ANOVA and the Tukey post-hoc test. P-values of 10-10 were deemed statistically significant. The Pearson correlation test was used to assess the relationship between the distribution of epialleles within each stage group. PCA was used to calculate the abundance of each of the 64 epialleles in the analyzed cell population. PC1 explained 31.2 percent of the observed variance, while PC2 explained 29 percent. JMP software was used for all statistical analyses (SAS, Cary, NC).

5. Conclusions

DNA methylation is an epigenetic modification essential for mammalian development and is crucial for establishing and maintaining cellular identity. Active DNA methylation and demethylation occur during cell fate commitment and terminal differentiation [39-40]. Recent data provide insights into the contribution of DNA methylation to the establishment of epigenetic memory during embryonic development and the modulation of cell type-specific gene regulatory programs to ensure proper differentiation [41]. Here, we demonstrated that the configurations of methyl-CpGs (methylated nuclei) define cellular identity and correlate with the regulation of gene expression.

Supplementary Materials: The following supporting information can be downloaded at the website of this paper posted on Preprints.org.

Author Contributions: M.C. and F.P. performed in vitro experiments. T.I., V.M., M.V., and A.F. performed bioinformatic analysis, analyzed the results, and prepared figures. F.P., L.C., M.C., A.Porcellini and A.Pezone supervised the experiments, analyzed the results, and provided scientific interpretations; A.Pezone wrote the paper. All authors discussed the results and provided input on the manuscript.

Funding: This work was supported by #NEXTGENERATIONEU (NGEU) and funded by the Ministry of University and Research (MUR), National Recovery and Resilience Plan (NRRP), project MNESYS (PE0000006) – A Multiscale integrated approach to the study of the nervous system in health and disease (DN. 1553 11.10.2022), A.Pezone is supported by the Italian Ministry of University and Scientific Research (P.R.I.N. 2017237P5x).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets generated for this study are available to the corresponding author upon request.

Acknowledgments: The authors thank Enrico Vittorio Avvedimento and Tommaso Russo for revising the manuscript and discussing dynamic epigenetics and cell identity.

Conflicts of Interest: The authors declare no conflict of interest.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 June 2023   doi:10.20944/preprints202305.2052.v2

Peer-reviewed version available at *Int. J. Mol. Sci.* **2023**, *24*, 9951; doi:https://doi.org/10.3390/ijms24129951

References

1. Moore, L., Le, T. & Fan, G. DNA Methylation and Its Basic Function. Neuropsychopharmacol 38, 23–38 (2013). https://doi.org/10.1038/npp.2012.112

2. Tramontano A., Boffo F.L., Russo G., De Rosa M., Iodice I., Pezone A. Methylation of the suppressor gene: mechanism and consequences. Biomolecules. 2020; 10:446. [Google Scholar]

3. Jin B, Ernst J, Tiedemann RL, et al. Linking DNA methyltransferases to epigenetic marks and nucleosome structure genome-wide in human tumor cells. Cell Rep. 2012;2(5):1411-1424. doi:10.1016/j.celrep.2012.10.017

4. Huang Y., Pastor W.A., Shen Y., Tahiliani M., Liu D.R., Rao A. The behavior of 5-hydroxymethylcytosine in bisulfite sequencing. PLoS One. 2010; 5:e8888. [PMC free article] [PubMed] [Google Scholar]

5. Rasmussen KD, Helin K. Role of TET enzymes in DNA methylation, development, and cancer. Genes Dev. 2016 Apr 1;30(7):733-50. doi: 10.1101/gad.276568.115. PMID: 27036965; PMCID: PMC4826392.

6. Huang Y, Pastor WA, Shen Y, Tahiliani M, Liu DR, Rao A. The behavior of 5-hydroxymethylcytosine in bisulfite sequencing. PLoS One. 2010;5(1):e8888. Published 2010 Jan 26. doi:10.1371/journal.pone.0008888

7. Song CX, He C. Potential functional roles of DNA demethylation intermediates. Trends Biochem Sci. 2013;38(10):480-484. doi:10.1016/j.tibs.2013.07.003

8. Xie H., Wang M., de Andrade A., Bonaldo M. de F., Galat V., Arndt K., Soares M.B. Genome-wide quantitative assessment of variation in DNA methylation patterns. Nucleic Acids Res. 2011; 39:4099–4108. [PMC free article] [PubMed] [Google Scholar]

9. Barros-Silva D, Marques CJ, Henrique R, Jerónimo C. Profiling DNA Methylation Based on Next-Generation Sequencing Approaches: New Insights and Clinical Applications. Genes (Basel). 2018 Aug 23;9(9):429. doi: 10.3390/genes9090429. PMID: 30142958; PMCID: PMC6162482.

10. Pezone A, Tramontano A, Scala G, et al. Tracing and tracking epiallele families in complex DNA populations. NAR Genom Bioinform. 2020;2(4):lqaa096. Published 2020 Nov 16. doi:10.1093/nargab/lqaa096

11. Podobinska M, Szablowska-Gadomska I, Augustyniak J, Sandvig I, Sandvig A, Buzanska L. Epigenetic Modulation of Stem Cells in Neurodevelopment: The Role of Methylation and Acetylation. Front Cell Neurosci. 2017;11:23. Published 2017 Feb 7. doi:10.3389/fncel.2017.00023

12. Kim M., Costello J. DNA methylation: an epigenetic mark of cellular memory. Exp. Mol. Med. 2017; 49:e322. [PMC free article] [PubMed] [Google Scholar]

13. Basu, A., Tiwari, V.K. Epigenetic reprogramming of cell identity: lessons from development for regenerative medicine. Clin Epigenet 13, 144 (2021). https://doi.org/10.1186/s13148-021-01131-4

14. Yang, Y.; Fan, X.; Yan, J.; Chen, M.; Zhu, M.; Tang, Y.; Liu, S.; Tang, Z. A comprehensive epigenome atlas reveals DNA methylation regulating skeletal muscle development. Nucleic Acids Res. 2021, 49, 1313–1329. [Google Scholar] [CrossRef]

15. Li S, Garrett-Bakelman F, Perl AE, et al. Dynamic evolution of clonal epialleles revealed by methclone. Genome Biol. 2014;15(9):472. Published 2014 Sep 27. doi:10.1186/s13059-014-0472-5

16. Li S, Garrett-Bakelman FE, Chung SS, et al. Distinct evolution and dynamics of epigenetic and genetic heterogeneity in acute myeloid leukemia. Nat Med. 2016;22(7):792-799. doi:10.1038/nm.4125

17. Shannon C.E. A mathematical theory of communication. Bell Syst. Tech. J. 1948; 27:379–423. [Google Scholar]

18. Verma R, Liu J, Holland MK, Temple-Smith P, Williamson M, Verma PJ. Nanog is an essential factor for induction of pluripotency in somatic cells from endangered felids. *Biores Open Access*. 2013;2(1):72-76. doi:10.1089/biores.2012.0297

19. Radwitz J, Hausrat TJ, Heisler FF, Janiesch PC, Pechmann Y, Rübhausen M, Kneussel M. Tubb3 expression levels are sensitive to neuronal activity changes and determine microtubule growth and kinesin-mediated transport. Cell Mol Life Sci. 2022 Oct 29;79(11):575. doi: 10.1007/s00018-022-04607-5. PMID: 36309617; PMCID: PMC9617967.

20. Errico F, Cuomo M, Canu N, Caputo V, Usiello A. New insights on the influence of free d-aspartate metabolism in the mammalian brain during prenatal and postnatal life. Biochim Biophys Acta Proteins Proteom. 2020;1868(10):140471. doi:10.1016/j.bbapap.2020.140471

21. Jurga AM, Paleczna M, Kadluczka J, Kuter KZ. Beyond the GFAP-Astrocyte Protein Markers in the Brain. Biomolecules. 2021 Sep 14;11(9):1361. doi: 10.3390/biom11091361. PMID: 34572572; PMCID: PMC8468264.

22. Bathina S, Das UN. Brain-derived neurotrophic factor and its clinical implications. Arch Med Sci. 2015 Dec 10;11(6):1164-78. doi: 10.5114/aoms.2015.56342. Epub 2015 Dec 11. PMID: 26788077; PMCID: PMC4697050.

23. Busskamp V, Lewis NE, Guye P, et al. Rapid neurogenesis through transcriptional activation in human stem cells. Mol Syst Biol. 2014;10(11):760. Published 2014 Nov 17. doi:10.15252/msb.20145508

24. Pang ZP, Yang N, Vierbuchen T, et al. Induction of human neuronal cells by defined transcription factors. Nature. 2011;476(7359):220-223. Published 2011 May 26. doi:10.1038/nature10202

25. Raspaglio G, Petrillo M, Martinelli E, et al. Sox9 and Hif-2α regulate TUBB3 gene expression and affect ovarian cancer aggressiveness. Gene. 2014;542(2):173-181. doi:10.1016/j.gene.2014.03.037

26. He M, Wang J. Decreased serum brain-derived neurotrophic factor in Chinese patients with Type 2 diabetes mellitus. Acta Biochim Biophys Sin (Shanghai). 2014;46(5):426-427. doi:10.1093/abbs/gmu008

27. Mamber C, Kamphuis W, Haring NL, Peprah N, Middeldorp J, Hol EM. GFAPδ expression in glia of the developmental and adolescent mouse brain. PLoS One. 2012;7(12):e52659. doi:10.1371/journal.pone.0052659

28. Ahmed AI, Shtaya AB, Zaben MJ, Owens EV, Kiecker C, Gray WP. Endogenous GFAP-positive neural stem/progenitor cells in the postnatal mouse cortex are activated following traumatic brain injury. J Neurotrauma. 2012;29(5):828-842. doi:10.1089/neu.2011.1923

29. Koga R., Miyoshi Y., Sakaue H., Hamase K., Konno R. Mouse D-amino-acid oxidase: distribution and physiological substrates. Front.Mol. Biosci.   2017; 4:82. [PMC free article] [PubMed] [Google Scholar]

30. Florio E, Keller S, Coretti L, et al. Tracking the evolution of epialleles during neural differentiation and brain development: D-Aspartate oxidase as a model gene. Epigenetics. 2017;12(1):41-54. doi:10.1080/15592294.2016.1260211

31. Punzo D, Errico F, Cristino L, et al. Age-Related Changes in D-Aspartate Oxidase Promoter Methylation Control Extracellular D-Aspartate Levels and Prevent Precocious Cell Death during Brain Aging. J Neurosci. 2016;36(10):3064-3078. doi:10.1523/JNEUROSCI.3881-15.2016

32. Breiling, A., Lyko, F. Epigenetic regulatory functions of DNA modifications: 5-methylcytosine and beyond. Epigenetics & Chromatin 8, 24 (2015). https://doi.org/10.1186/s13072-015-0016-6

33. Luo C, Hajkova P, Ecker JR. Dynamic DNA methylation: In the right place at the right time. Science. 2018;361(6409):1336-1340. doi:10.1126/science.aat6806

34. Fuentealba L.C., Rompani S.B., Parraguez J.I., Obernier K., Romero R., Cepko C.L., Alvarez-Buylla A. Embryonic origin of postnatal neural stem cells. Cell. 2015; 161:1644–1655. [PMC free article] [PubMed] [Google Scholar]

35. Anderson RB, Newgreen DF, Young HM. Neural Crest and the Development of the Enteric Nervous System. In: Madame Curie Bioscience Database [Internet]. Austin (TX): Landes Bioscience; 2000-2013. Available from: https://www.ncbi.nlm.nih.gov/books/NBK6273/

36. Niwa, H., Miyazaki, Ji. & Smith, A. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. Nat Genet 24, 372–376 (2000). https://doi.org/10.1038/74199.

37. Parisi S, Passaro F, Aloia L, Manabe I, Nagai R, Pastore L, Russo T. Klf5 is involved in self-renewal of mouse embryonic stem cells. J Cell Sci. 2008, 121(Pt 16):2629-34. https://doi.org/10.1242/jcs.027599

38. Passaro F, De Martino I, Zambelli F, Di Benedetto G, Barbato M, D'Erchia AM, Manzari C, Pesole G, Mutarelli M, Cac-chiarelli D, Antonini D, Parisi S, Russo T. YAP contributes to DNA methylation remodeling upon mouse embryonic stem cell differentiation. J Biol Chem. 2021, 100138. https://doi.org/10.1074/jbc.RA120.015896

39. Russo G, Tramontano A, Iodice I, Chiariotti L, Pezone A. Epigenome Chaos: Stochastic and Deterministic DNA Methylation Events Drive Cancer Evolution. Cancers (Basel). 2021;13(8):1800. Published 2021 Apr 9. doi:10.3390/cancers13081800

40. Pezone A, Russo G, Tramontano A, et al. High-coverage methylation data of a gene model before and after DNA damage and homologous repair. Sci Data. 2017;4:170043. Published 2017 Apr 11. doi:10.1038/sdata.2017.43

41. Suelves M, Carrió E, Núñez-Álvarez Y, Peinado MA. DNA methylation dynamics in cellular commitment and differentiation. Brief Funct Genomics. 2016;15(6):443-453. doi:10.1093/bfgp/elw017.