Intro

**What is epigenetics?**

Epigenetics is the study of heritable and reversible changes of the genome that do not change the underlying sequence. These changes are fundamental for controlling gene expression in all organisms, and can occur through intrinsic methods (e.g., X-chromosome inactivation in females) or external causes (e.g., exposure to ultra-violet radiation). The molecular marks of epigenetics span several levels from overall chromatin structure, through histone modifications, down to changes of individual bases, with methylation.

**Describe epigenetic marks**

Methylation is the most studied epigenetic modification. Adenine and cytosine both readily accept methylation on N4and C4/C5, respectively. Methylated adenine is weakly studied but widely known in bacteria and lower eukaryotes, but is not considered to have any effect or significant presence in mammalian DNA1,2. Comparatively, cytosine modifications are abundant and well-studied among all domains of life. 5-methyl-cytosine (5mC) is the most common epigenetic mark in humans, with about 28 million sites present in the genome3. These sites, known as CpG sites for the cytosine-phosphate-guanine bridge, are not randomly distributed but instead highly localized around regulatory motifs: 70% of promoters show high density of CpG sites4. This density is called a CpG island (CGI). The bulky methyl groups at CpG sites provide a physical block to binding of transcription factors or even RNA polymerase itself5. As such, methylation is a major factor in gene expression in mammals, so mechanisms for adding and removing methylation must be strictly regulated.

These epigenetic marks can be *de novo* added by DNA Methyltransferase proteins (DNMT3a/b), or removed by ten-eleven translocation methylcytosine dioxygenases (TET family), so these modifications are considered reversible. These are necessary for normal development of tissues, as Example of adding and removing. Furthermore, these modifications can also be maintained through cell division by DNMT1, hence they are considered heritable. Example of heritable

These mechanisms must be strictly controlled, as dysregulation can significantly affect health and disease. Example of sickness by epigenetics

Implications to cancer

Malignant cell growth is also strongly linked to epigenetic abnormalities. Abnormal overexpression of DNMT proteins has shown significant effects on multiple cancers (e.g. DNMT3A in 25% of acute myeloid leukemia cases and DNMT1 in 12% of uterine cancer cases <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7465608/#:~:text=found%20that%20the%20coding%20exons,characteristic%20change%20of%20tumor%20cells>.). Conversely, underexpression of TET proteins has been shown in some cancer types (e.g. up to 58% of chronic myelomonocytic leukemia cases). Thus, epimutations can both increase and decrease gene expression, which is problematic with pro-tumor and tumor-suppressing genes, respectively. Example related to later work. Alterations of the epigenomic landscape can cause widespread tumor and disease, and understanding this process is critical for well being.

Whole genome bisulfite sequencing (WGBS) is the *de facto* technique for studying the epigenomic landscape. In this technique, sodium bisulfite converts unmethylated cytosines to uracil (and later thymine) which allows differentiation by comparing pre- and post-treatment sequencing, then aligning to the reference methylome. The resulting single nucleotide polymorphisms (SNPs) allow generation of the epigenomic landscape. A coverage of at least 30x is recommended with this technique, as during treatment, single strand nicks are randomly introduced, so up to 95% of the CpG sites lost during sequencing cite.

This causes problems with low population samples. Recently, new techniques focusing on single cell data have emerged to address these problems. Single cell bisulfite sequencing (sc-BS) and XXX

Existing methods

Importance of single cell methods

However, WGBS suffers some drawbacks due to its bulk processing. Low population samples cannot give suitable coverage and is difficult to differentiate between heterogenous cells. Single cell bisulfite sequencing (sc-BS) fills this gap.

Furthermore, the data obtained from sc-BS presents its own problems. While low coverage can be accommodated by many tools, sparsity of the data is a challenge. There are inherent losses with bisulfite treatment as described above. Existing WGBS tools are unable of handle a high proportion of NA values, and there is a high memory requirement if stored in a dense matrix, as most also require. Hence, we have developed a new tool to address these problems and allow a pipeline more similar to traditional WGBS tools.

To address these problems, we developed scMethrix, a methylation-specific single cell data storage and manipulation tool. This R package is well-suited for the specific needs of single cell data and for integration into the Bioconductor ecosystem.

https://www.nature.com/articles/nri.2017.125

Ly6Chi classical inflammatory monocytes (CD14hiCD16− in humans) and Ly6Clow non-classical patrolling monocytes (CD14+CD16hi in humans)[35](https://www.nature.com/articles/nri.2017.125#ref-CR35). Ly6Clow monocytes are derived from Ly6Chi monocytes in either the blood or the bone marrow (BM)[36](https://www.nature.com/articles/nri.2017.125#ref-CR36). During certain diseases or injuries involving breach of the [blood–brain barrier](https://www.nature.com/articles/nri.2017.125#Glos1) (BBB), Ly6Chi monocytes may infiltrate the brain parenchyma and differentiate into microglia-like cells, which are intermingled with the resident microglia, to exacerbate or alleviate disease progression[29](https://www.nature.com/articles/nri.2017.125#ref-CR29),[35](https://www.nature.com/articles/nri.2017.125#ref-CR35).

Glioma has shown strong correlation with macrophages such as native microglia and infiltrating monocytes6. These tumor-associated macrophages (TAMs) have previously shown promise as a target for treatment7.

Using epigenetic lineage, we hypothesize that methylation can be used to identify and help determine the origins of TAMs implicated in gliomagenesis. With this knowledge, it may be possible to generate prognostic outcomes based on TAM presence, improve patient-specific treatment, and may aid in finding further avenues of research for fighting this devastating disease.

* Origins of brain monocytes
  + Monocytes vs microglia evolution

Methods

Raw data files were obtained from publicly available sources, including monocyte, neutrophil, NK cells, eosinophils from whole blood samples of 44 individuals (GSE35069, GSE88824, GSE166844), glia (GSE66351), microglia-like cells from 25 mouse samples (GSE121483), glioma-related immune cells from 77 individuals (GSE151506).

Methylation array data (e.g., Illumina 450k and EPIC) were converted to genomic positions via the Minfi R package8. All genome assemblies were translated to hg38 by the Rtracklayer::liftOver R package9. For data sourced from mouse models (GSE121483), only the 19 420 probes conserved with humans were kept10, and it has been show that CpG methylation for glial cells is highly conserved between mice and humans11.

Scripts for data importing are available on Github (https://github.com/knacko/monobrainDNAme).

1. Bochtler, M. & Fernandes, H. DNA adenine methylation in eukaryotes: Enzymatic mark or a form of DNA damage? *BioEssays* **43**, 2000243 (2021).

2. Douvlataniotis, K., Bensberg, M., Lentini, A., Gylemo, B. & Nestor, C. E. No evidence for DNA N6-methyladenine in mammals. *Science Advances* **6**, eaay3335.

3. Bird, A. P. DNA methylation and the frequency of CpG in animal DNA. *Nucleic Acids Res* **8**, 1499–1504 (1980).

4. Saxonov, S., Berg, P. & Brutlag, D. L. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc Natl Acad Sci U S A* **103**, 1412–1417 (2006).

5. Deaton, A. M. & Bird, A. CpG islands and the regulation of transcription. *Genes Dev* **25**, 1010–1022 (2011).

6. Zhang, N. *et al.* The Predictive Value of Monocytes in Immune Microenvironment and Prognosis of Glioma Patients Based on Machine Learning. *Front Immunol* **12**, 656541 (2021).

7. Grégoire, H. *et al.* Targeting Tumor Associated Macrophages to Overcome Conventional Treatment Resistance in Glioblastoma. *Front Pharmacol* **11**, 368 (2020).

8. Aryee, M. J. *et al.* Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* **30**, 1363–1369 (2014).

9. Bioconductor Package Maintainer. liftOver: Changing genomic coordinate systems with rtracklayer::liftOver. R package version 1.18.0. https://www.bioconductor.org/help/workflows/liftOver/.

10. Needhamsen, M. *et al.* Usability of human Infinium MethylationEPIC BeadChip for mouse DNA methylation studies. *BMC Bioinformatics* **18**, 486 (2017).

11. Kessler, N. J. *et al.* CpG methylation differences between neurons and glia are highly conserved from mouse to human. *Human Molecular Genetics* **25**, 223–232 (2016).