SCIENTIFIC DATA

OPEN

Data Descriptor: Brain metastasis DNA methylomes, a novel resource for the identification of biological and clinical features

Matthew P. Salomon[1,*], Javier I. J. Orozco[1,*], James S. Wilmott[2], Parvinder Hothi[3], Ayla O. Manughian-Peter[1], Charles S. Cobbs[3], Richard A. Scolyer[2,4,5], Dave S. B. Hoon[1] & Diego M. Marzese[1]

Brain metastases (BM) are one the most lethal and poorly managed clinical complications in cancer patients. These secondary tumors represent the most common intracranial neoplasm in adults, most frequently originating from lung cancer, breast cancer, and cutaneous melanoma. In primary brain tumors, such as gliomas, recent advances in DNA methylation profiling have allowed for a comprehensive molecular classification. Such data provide prognostic information, in addition to helping predict patient response to specific systemic therapies. However, epigenetic alterations of metastatic brain tumors with specific biological and translational relevance still require much further exploration. Using the widely employed Illumina Infinium HumanMethylation 450K platform, we have generated a cohort of genome-wide DNA methylomes from ninety-six needle-dissected BM specimens from patients with lung cancer, breast cancer, and cutaneous melanoma with clinical, pathological, and demographic annotations. This resource offers an unprecedented and unique opportunity to identify novel DNA methylation features influencing the behavior of brain metastasis, and thus accelerate the discovery of BM-specific theranostic epigenetic alterations.

Design Type(s)	disease state design • DNA methylation profiling by array design
Measurement Type(s)	DNA residue methylation
Technology Type(s)	DNA methylation profiling assay
Factor Type(s)	Primary Malignant Neoplasm • Tumor Subtype
Sample Characteristic(s)	Homo sapiens • brain

[1]Department of Translational Molecular Medicine, John Wayne Cancer Institute at Providence Saint John's Health Center, Santa Monica, CA 90404, USA. [2]Melanoma Institute Australia, The University of Sydney, Camperdown, NSW, 2065, Australia. [3]Ben & Catherine Ivy Center for Advanced Brain Tumor Treatment, Swedish Neuroscience Institute, Seattle, WA 98122, USA. [4]Sydney Medical School, The University of Sydney, Camperdown, NSW, 2006, Australia. [5]Royal Prince Alfred Hospital, Sydney, New South Wales, 2050, Australia. *These authors contributed equally to this work. Correspondence and requests for materials should be addressed to D.M.M. (email: marzesed@jwci.org)

Background & Summary

Patients diagnosed with brain metastasis (BM) have a poor quality of life and a dismal prognosis, with survival ranging from three to 25 months[1,2]. While developments in systemic drug treatments have significantly improved the survival of patients with extracranial metastases, BM lesions have shown limited response rates to these approaches[3]. This trend, along with improvements in neuro-diagnostic imaging techniques, has resulted in an increased incidence of metastatic brain tumors[4]. In large epidemiological studies, 8–10% of patients with solid tumors develop BM, this number increases up to 26% when brain autopsy studies are performed[5–9]. In fact, BM represents the most common intracranial neoplasm in adults, outnumbering even primary brain tumors. Interestingly, reflecting an organotropic behavior of tumor cells, the vast majority of secondary brain neoplasms (75-90%) are originated in patients with lung cancer, breast cancer, and cutaneous melanoma[5–8].

Genomic and epigenomic landscapes of primary brain tumors have been extensively investigated. In gliomas, for example, the cytosine-guanine island (CGI) methylator phenotype (CIMP) is frequently found in patients with lower grade gliomas harboring mutations on the *IDH1* gene and is significantly associated with a better overall prognosis[10]. Additially, a favorable response to the DNA alkylating antineoplastic agent, temozolomide, has been directly connected to high DNA methylation (DNAm) level in the promoter region of the *MGMT* gene[11]. DNAm profiling has recently been shown to have the potential to accurately stratify primary central nervous system (CNS) tumours[12] and to significantly improve the diagnosis of cancer of unknown primary[13]. These clinically relevant findings have demonstrated DNAm profiling to be a valuable tool in the histomolecular evaluation of brain tumors[14,15]. Yet, while genomic and transcriptomic characterization has been performed to some extent[16], clinically relevant epigenetic alterations of metastatic brain tumors are still poorly understood. Therefore, given this significant knowledge gap, we constructed a comprehensive dataset that can be used to accelerate the identification of novel DNAm features with biological and clinical relevance for the three most frequent types of BM. Here, we present a dataset including genome-wide DNA methylomes constructed using Illumina Infinium HumanMethylation 450K BeadChips (HM450K) of 96 micro-dissected BM specimens from patients with breast cancer, lung cancer, and cutaneous melanoma (Fig. 1). In addition to DNAm data, this report provides a detailed description of the methodological approaches for patient selection, compliance matters, tissue processing and DNA preparation, data normalization, bioinformatics analyses, and usage notes including clinical and demographic information for all patients in the study. Seven of these patients are part of a cohort study that we previously analyzed to identify genome-wide DNAm variations during cutaneous melanoma progression to BM[17–20]. Therefore, the current cohort of BM DNA methylomes is composed of HM450K profiles included in two different NCBI's Gene Expression Omnibus (GEO) datasets (GSE108576 and GSE44661). We believe that these datasets offer a unique opportunity for the discovery of novel diagnostic and prognostic biomarkers, while simultaneously providing insight into the underlying biology of this serious clinical complication. In this regard, we have employed these data to further explore the utility of DNAm profiles to accurately discriminate between primary and metastatic brain tumors, identify the origin of the BM lesions, and specifically classify BCBM into therapeutically relevant molecular subtypes[21]. Thus, we generated and validated a three-steps BM DNAm based classifier named "BrainMETH"[21].

Methods

Tissue specimen collection

A total of 96 metastatic brain formalin-fixed paraffin-embedded (FFPE) tumor samples from 94 patients diagnosed with breast cancer BM (BCBM; n = 30), lung cancer BM (LCBM; n = 22), and cutaneous melanoma BMs (MBM; n = 44) were included in this study. Two breast cancer patients presented synchronous or asynchronous multiple lesions. The clinical and demographic characteristics of the patients included in the study have been summarized according to relevant information for each cancer type (Table 1). All patient-derived samples and clinical and demographic data were collected under research protocols approved by the joint Institutional Review Board of Providence Saint John's Health Center/John Wayne Cancer Institute, the Western Institutional Review Board, the Institutional Review Board of Swedish Medical Center, and the Sydney Local Health District (Royal Prince Alfred Hospital Zone) Human Ethics Review Committee. All patients signed an informed consent before joining the study. The experiments were performed in accordance with the World Medical Association Declaration of Helsinki and the National Institutes of Health Belmont Report. Tissues were de-identified and coded according to recommendations of the Health Insurance Portability and Accountability Act (HIPAA) to ensure confidentiality of the patients.

Histopathological classification of brain metastasis

The BCBM specimens were classified into molecular subtypes according to the expression status of the hormone receptors (HR), i.e. estrogen receptor (ER) and progesterone receptor (PgR), and the human epidermal growth factor receptor 2 (HER2). ER and PgR were assessed by immunohistochemistry (IHC), and HER2 by IHC and/or *in situ* hybridization assays (ISH). FFPE tissue slides were sectioned at 4 μm, mounted onto plus-coated glass slides, and immunohistochemically stained using a Ventana BenchMark ULTRA automated slide stainer (Roche Diagnostics, Indianapolis, IN, USA) by the Clinical Laboratory Improvement Amendments (CLIA)-certified Department of Pathology, Providence Saint John's Health
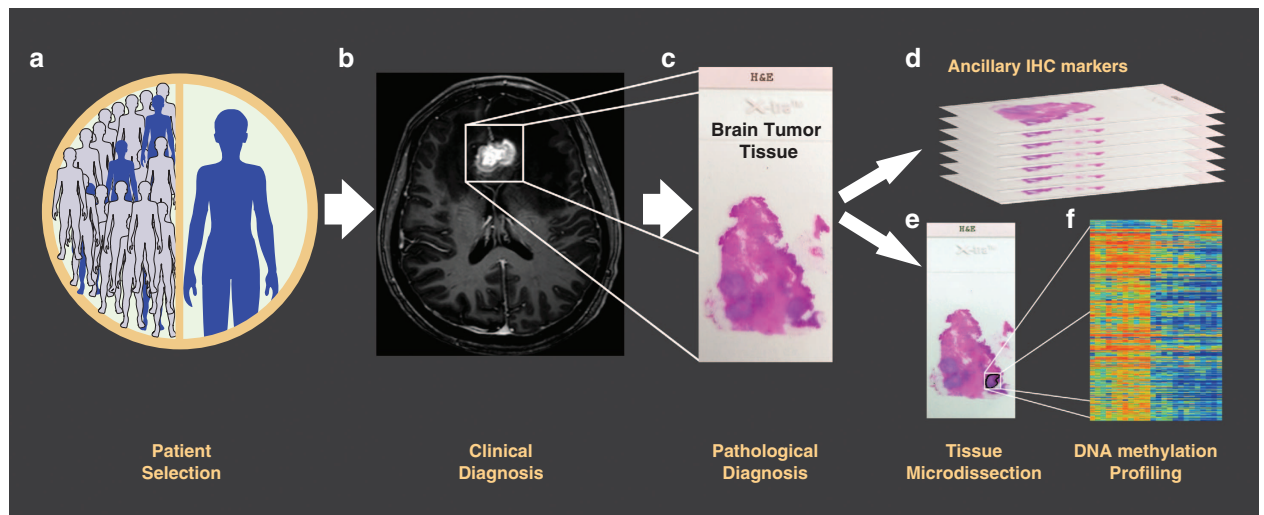
Figure 1. Study design for the construction of genome-wide DNA methylation profiles of metastatic brain tumors. (**a**) Patients with metastatic brain tumors from breast cancer, lung cancer or cutaneous melanoma origin were selected for the study. (**b**) A representative magnetic resonance imaging (MRI) scan of a single metastatic brain tumor lesion used as part of the clinical diagnosis is shown in the scheme. (**c**) After surgery, resected tumors were routinely stored as formalin-fixed and paraffin-embedded (FFPE) tissue blocks and stained with hematoxylin and eosin (H&E) for anatomic pathology diagnosis. (**d**) FFPE tissue sections underwent routine immunohistochemistry (IHC) evaluation to confirm the tumor of origin and molecular subtypes of each case. (**e**) After tumor cell-rich areas were identified, tissue microdissection followed by DNA purification was performed in each case. (**f**) DNA specimens passing the quality control metrics were converted with sodium bisulfite, enzymatically fragmented, and hybridized in the HM450K BeadChips. Raw intensity data were normalized and corrected β values for each specimen were generated. A representative heat map of the DNA methylation data is shown in the study scheme.

	BCBM	LCBM	MBM	Total BM
N	28	22	44	94
Age (years)				
Median	54.5	67.5	58.45	61.35
IQR	19.75	10	20.90	19.81
Range	35-79	43-88	32-82	32-88
Gender				
Female	28 (100%)	18 (81.8%)	12 (27.27%)	58 (61.7%)
Male	0	4 (18.2%)	32 (72.73%)	36 (38.3%)
Cancer Specific Subtypes				
	HR+/HER2-: 12 (42.86%)	NSCLC: 19 (86.4%)	BRAF mut: 19 (43.19%)	
	HR any/HER2+: 11 (39.28%)	SCLC: 3 (13.6%)	NRAS mut: 18 (40.91%)	
	HR-/HER2-: 5 (17.86%)		BRAF/NRAS wt: 5 (11.36%)	
			N/A: 2 (4.54%)	

Table 1. Clinical-demographic characteristics of patients with brain metastasis.

Center, accredited by the College of American Pathologists (CAP). The antibodies used in this evaluation were the CONFIRM anti-Estrogen Receptor (SP1, #790-4324, Ventana Medical Systems, Tucson, AZ, USA), the CONFIRM anti-Progesterone Receptor (1E2, #790-2223, Ventana Medical Systems, Tucson, AZ, USA), and the PATHWAY anti-HER-2/neu (4B5, #790-2991, Ventana Medical Systems, Tucson, AZ, USA). The scoring criteria for these biomarkers were based on the current ASCO/CAP guidelines[22,23]. Briefly, ER and PgR were considered positive if there was staining of the nucleus in at
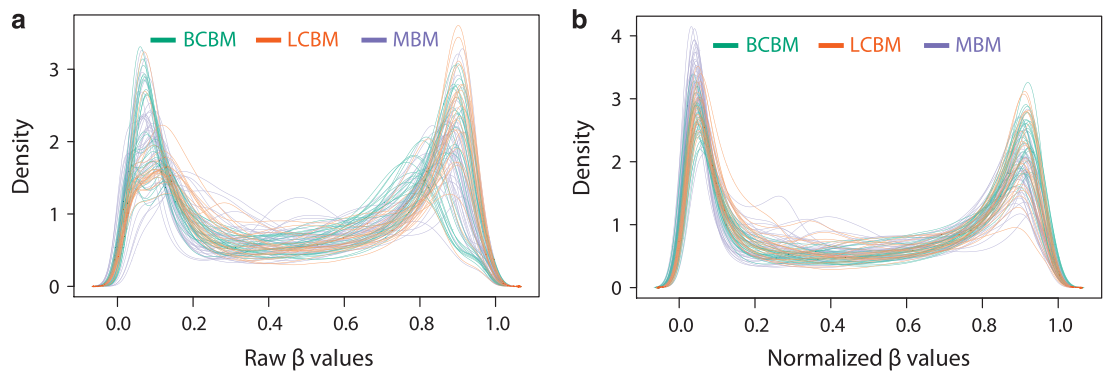
Figure 2. **Density distribution of β values across samples.** (**a**) The distribution of raw β values for all three BM subtypes before normalization. (**b**) The distribution of β values after normalization for all three BM subtypes. BM subtype is indicated by the color of the line, with green lines representing breast cancer brain metastases (BCBM), orange lines representing lung cancer brain metastases (LCBM), and purple lines representing melanoma brain metastases (MBM).

least ≥1% of tumor cells in the sample. HER2 test result was considered positive if IHC 3+ (observed in a homogeneous and contiguous population and within >10% of the invasive tumor cells) or ISH amplified if single-probe average HER2 copy number >6.0 signals/cell or dual-probe HER2/CEP17 ratio ≥2.0. BCBM specimens were grouped according to the expression of these routinely evaluated markers into three therapeutically relevant subgroups: a- HR positive/HER2 negative, b- HR any/HER2 positive (HER2+), and c- HR negative/HER2 negative (aka triple-negative breast cancer; Table 1).

The MBM samples were categorized according to the mutational status of *BRAF* and *NRAS* genes. Genomic DNA from MBM was amplified with standardized primers specific for exon 15 of *BRAF*, and exons 1 and 2 of *NRAS*[20]. Polymerase chain reaction (PCR) products were purified using QIAquick® PCR Purification Kit (#28106 Qiagen, Germany) and subsequently visualized in 2.2% agarose gel DNA cassettes for gel electrophoresis (FlashGel™ System, Lonza Inc, Rockland, ME, USA). Successfully amplified samples were then quantified by UV absorption spectrophotometry and sequenced using an internal primer[20] by Eurofins MWG Operon LCC (Eurofins Genomics LCC, Louisville, KY, USA). Sequencing results were analyzed using Chromas Lite v2.6.5 (Technelysium Pty Ltd, Australia) and mutations in *NRAS* and *BRAF* genes were annotated according to the Catalogue of Somatic Mutations in Cancer (COSMIC v86, Wellcome Sanger Institute, Cambridge, UK; http://cancer.sanger.ac.uk/cosmic)[24]. As the presence of *BRAF* and *NRAS* mutations were mutually exclusive events, the MBM specimens were classified into 3 categories: a- BRAF mutated, b- NRAS mutated, and c- BRAF/NRAS wild-type (Table 1). Due to limited tissue availability, two specimens were not profiled for oncogenic mutations on the *BRAF* or *NRAS* genes and presented in Tables 1 and 2 as not available (N/A).

The LCBM were histologically classified into non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC; Table 1). Of note, we added four BM specimens from female patients with a presumptive diagnosis of LCBM, but with inconclusive IHC analysis, or with a previous diagnosis of both primary lung and breast cancer. In agreement with the clinical-pathological diagnosis an origin of lung cancer was confirmed by DNAm profiling[21].

Genomic DNA extraction

Representative FFPE tissue blocks from each metastatic brain lesion were selected by the respective Pathology Departments. FFPE tissue blocks were cut into 4 μm and 8 μm serial slides. Neuropathologists reviewed 4 μm tissue slides stained with hematoxylin & eosin (H&E) for all specimens and labeled representative brain metastatic areas with an estimated tumor purity higher than 70%. After deparaffinization, hematoxylin staining was performed in 8 μm thick serial tissue sections and needle microdissected using the labeled 4 μm tissue slides as template. Genomic DNA (gDNA) was then isolated using ZR FFPE DNA MiniPrep (D3066; Zymo Research, Irvine, CA, USA), according to the manufacturer's instructions. Genomic DNA was quantified by Qubit® 3.0 Fluorometer (Q33216; Thermo Fisher Scientific, Carlsbad, CA).

Genome-wide DNA methylation profiling

Sodium bisulfite modification (SBM) was performed on 1 μg of gDNA using the EZ DNA Methylation-Direct Kit (D5021, Zymo Research Irvine, CA, USA). An aliquot of SBM-DNA was analyzed by MethyLight-based quality control to test bisulfite completeness. After correction of SBM-DNA amount, a minimum of 200 ng of SBM-DNA was whole-genome amplified and enzymatically fragmented. Finally, the fragmented SBM-DNA was hybridized into the HM450K (Illumina Inc., San Diego, CA, USA) and
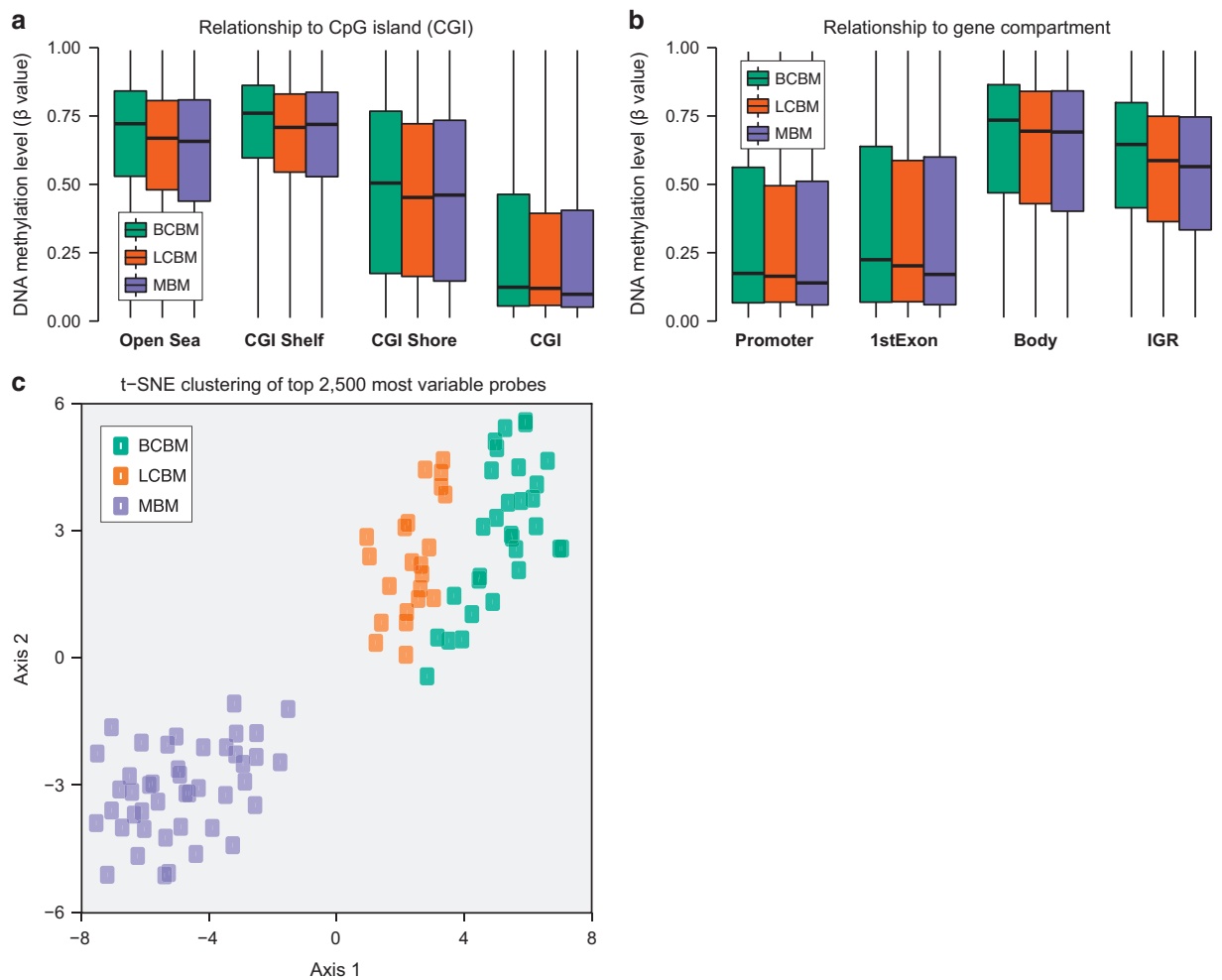
Figure 3. Distribution of β values among genomic features and BM type specific clustering. (**a**) The distribution of DNAm levels in low-CpG density regions (i.e. open sea, CGI shelf, and CGI shore) with respect to high-CpG density CGIs. (**b**) The distribution of DNAm levels for CpG sites with respect to gene compartments in the promoter regions, 5'UTRs, the 1st exon, gene body, and intergenic regions (IGRs). Box colors indicate BM subtype, with green boxes representing breast cancer brain metastases (BCBM), orange boxes representing lung cancer brain metastases (LCBM), and purple boxes representing melanoma brain metastases (MBM). (**c**) t-SNE analysis reveals BM subtype specific clustering of samples. The top 2,500 most variable CpG sites were used for the analysis with the first two dimensions shown. Each of the three BM subtypes is confined to a cluster, with MBMs depicted in purple, BCBMs depicted in green, and LCBMs depicted in orange. No samples were found to fall outside of a BM subtype specific cluster

scanned using the Illumina iScan microarray scanner following the manufacturer's recommended settings (Illumina Inc., San Diego, CA, USA).

Data processing and analysis
Data was extracted from Illumina .idat files using the Bioconductor package *minfi*[25]. The '*preprocessNoob*' function in *minfi* was used for normalization and dye-bias correction as described in Triche *et al.*[26]. DNAm levels were reported as β-values [β = intensity of the methylated allele/(intensity of the unmethylated allele + intensity of the methylated allele)], and calculated using the signal intensity value for each CpG site. The effect of normalization on the distribution of β values across samples is shown in Fig. 2.

Using the normalized β values, we compared the genome-wide DNAm profiles for specific genomic features across the three BM groups. DNAm level of CpG sites in high-CpG density regions (known as CpG islands; CGI) and low-CpG density regions (known as CGI shore, CGI shelves, and open sea) were also variable among the three BM groups (Fig. 3a). Additionally, DNAm levels varied among the three BM groups for CpG sites in the promoter regions, 5'UTRs, the first exon, gene body, and intergenic
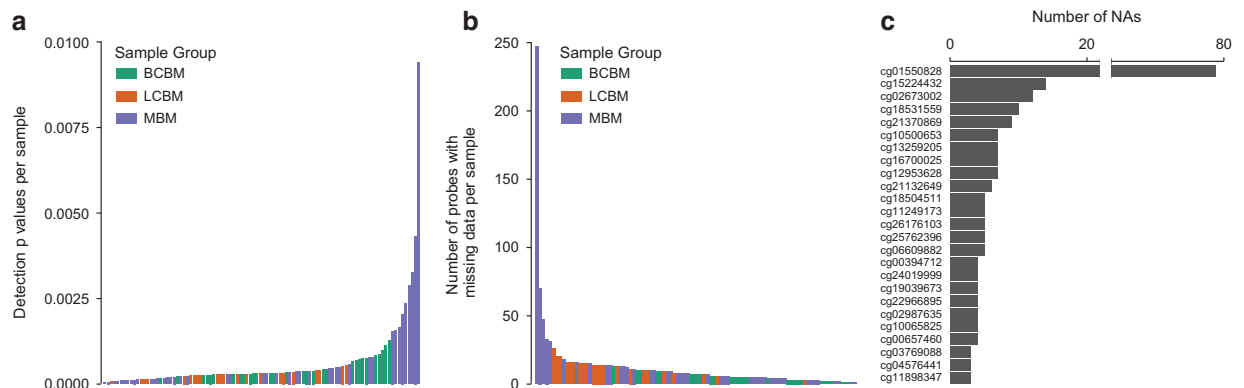
Figure 4. **Quality metrics used for sample filtering. A three-part quality control pipeline was used to identify samples with low quality.** (**a**) The distribution of mean detection *p*-values for all probes in each sample. (**b**) The distribution of the number of probes with missing data per sample. (**c**) The distribution of the top 25 probes with missing data. For panels **a** and **b**, colors represented BM subtype with green bars representing breast cancer brain metastases (BCBM), orange bars representing lung cancer brain metastases (LCBM), and purple bars representing melanoma brain metastases (MBM).

regions (IGRs; Fig. 3b). Finally, to check for overall structure within our dataset, we used the t-distributed stochastic neighbor embedding (t-SNE)[27,28] method with the top 2,500 most variable HM450K probes to cluster all BM specimens. Three distinct clusters were observed that corresponded to each of the three BM types, with MBMs showing the greatest degree of separation from BCBM and LCBM which were positioned more closely to each other (Fig. 3c). No outlier samples were observed.

Code availability

All analyses were performed using open source R and Bioconductor packages. Specifically, the *minfi*[25] package was used to process raw array data and perform normalizations (see "Data processing and analysis" section), summary statistics were calculated using functions in base R and the *matrixStats*[29] package, density distribution plots were generated using the *densityPlot* function in *minfi*[25], all other figures were generated using the *ggplot2*[30] and *RColorBrewer*[31] packages, and the t-SNE analysis was performed using the *Rtsne*[32] package. No custom code was used in the processing or analysis of this data.

Data Records

All HM450K raw and normalized data that support the findings of this study have been deposited in the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/geo/) datasets under the series records GSE108576 (Data Citation 1) and GSE44661 (Data Citation 2). The data is presented in a tabular format that includes the unmethylated intensity values, the methylated intensity values, the *p*-value from the statistical evaluation of the differences between signal and noise, and the corrected β value. The DNAm data can also be accessed as raw intensity files (.idat). Additionally, the integration of the clinical-demographic characteristics of the 96 BM specimens with the matched .idat file names and the GEO sample identifiers (GSM) is provided in Table 2 (available online only).

Technical Validation

To ensure that only samples with high overall quality were included in this dataset, we applied a three-step quality control pipeline: 1) We filtered samples by probe detection *p*-value to identify samples with an elevated level of background noise. A significance level of 0.05 for the mean per sample detection *p*-value was used as a cut-off. All 96 samples included in this dataset showed mean detection *p*-values less than 0.05 (Fig. 4a). 2) We calculated the number of probes with missing β values per sample. Across all 96 samples, the median number of probes with missing β values was 7.0 probes per sample with a range of 1 to 246 probes (Fig. 4b). Overall, the number of probes with missing β values represents a minuscule fraction of the total number of probes present on the array and therefore is highly unlikely to have an adverse effect on downstream analysis. 3) For each probe, we calculated the number of samples with missing β values. Notably, of the 485,577 probes included on the HM450K microarray, probe cg01550828 showed missing β values in 79 samples (Fig. 4c). Probe cg01550828 is located in the body of the ring finger protein 168 (*RNF168*) gene and is one of five probes within the *RNF168* gene body. While cg01550828 showed missing values, none of the other four *RNF168* gene body probes showed any missing values across the 96 samples.

Usage Notes

To enhance the utility of this resource, we have integrated the most relevant clinical and demographic features of the patient cohort and DNAm data for each BM specimen. In Table 2 (available online only), we included patient age at BM diagnosis, gender, primary cancer of origin, and cancer-specific subtypes matched with GEO sample names and .idat identifiers. This information can be accessed from the respective GEO series GSE108576 (Data Citation 1) and GSE44661 (Data Citation 2).

The dataset we present here can be further analyzed to study the differential methylation profiles among the three BM groups described here and/or integrated into larger methylation analyses using new or existing publicly available array data deposited in GEO. Data normalization and differential methylation analysis can be performed using various open source Bioconductor packages. In particular, the ChAMP Bioconductor package provides a comprehensive analysis pipeline that utilizes many well-established methods for the normalization and analysis of Illumina HM450K microarray data[33]. This package is well documented and provides a useful first pass pipeline for processing array data.

References

1. Sperduto, P. W. *et al.* Summary report on the graded prognostic assessment: an accurate and facile diagnosis-specific tool to estimate survival for patients with brain metastases. *J. Clin. Oncol.* **30,** 419–425 (2012).
2. Soffietti, R. *et al.* A European Organisation for Research and Treatment of Cancer phase III trial of adjuvant whole-brain radiotherapy versus observation in patients with one to three brain metastases from solid tumors after surgical resection or radiosurgery: quality-of-life results. *J. Clin. Oncol.* **31,** 65–72 (2013).
3. Lin, N. U., Bellon, J. R. & Winer, E. P. CNS metastases in breast cancer. *J. Clin. Oncol.* **22,** 3608–3617 (2004).
4. Lin, X. & DeAngelis, L. M. Treatment of Brain Metastases. *J. Clin. Oncol.* **33,** 3475–3484 (2015).
5. Schouten, L. J., Rutten, J., Huveneers, H. A. & Twijnstra, A. Incidence of brain metastases in a cohort of patients with carcinoma of the breast, colon, kidney, and lung and melanoma. *Cancer* **94,** 2698–2705 (2002).
6. Barnholtz-Sloan, J. S. *et al.* Incidence proportions of brain metastases in patients diagnosed (1973 to 2001) in the Metropolitan Detroit Cancer Surveillance System. *J. Clin. Oncol.* **22,** 2865–2872 (2004).
7. Berghoff, A. S. *et al.* Descriptive statistical analysis of a real life cohort of 2419 patients with brain metastases of solid cancers. *ESMO Open* **1,** e000024 (2016).
8. Cagney, D. N. *et al.* Incidence and prognosis of patients with brain metastases at diagnosis of systemic malignancy: A population-based study. *Neuro Oncol.* (2017).
9. Gavrilovic, I. T. & Posner, J. B. Brain metastases: epidemiology and pathophysiology. *J. Neurooncol.* **75,** 5–14 (2005).
10. Noushmehr, H. *et al.* Identification of a CpG island methylator phenotype that defines a distinct subgroup of glioma. *Cancer Cell* **17,** 510–522 (2010).
11. Hegi, M. E. *et al.* MGMT gene silencing and benefit from temozolomide in glioblastoma. *N. Engl. J. Med.* **352,** 997–1003 (2005).
12. Capper, D. *et al.* DNA methylation-based classification of central nervous system tumours. *Nature* **555,** 469 (2018).
13. Moran, S. *et al.* Epigenetic profiling to classify cancer of unknown primary: a multicentre, retrospective analysis. *Lancet Oncol.* **17,** 1386–1395 (2016).
14. Ceccarelli, M. *et al.* Molecular Profiling Reveals Biologically Discrete Subsets and Pathways of Progression in Diffuse Glioma. *Cell* **164,** 550–563 (2016).
15. Sahm, F. *et al.* DNA methylation-based classification and grading system for meningioma: a multicentre, retrospective analysis. *Lancet Oncol.* **18,** 682–694 (2017).
16. Brastianos, P. K. *et al.* Genomic Characterization of Brain Metastases Reveals Branched Evolution and Potential Therapeutic Targets. *Cancer Discov* **5,** 1164–1177 (2015).
17. Marzese, D. M. *et al.* Epigenome-wide DNA methylation landscape of melanoma progression to brain metastasis reveals aberrations on homeobox D cluster associated with prognosis. *Hum. Mol. Genet* **23,** 226–238 (2014).
18. Marzese, D. M., Huynh, J. L., Kawas, N. P. & Hoon, D. S. Multi-platform Genome-wide Analysis of Melanoma Progression to Brain Metastasis. *Genom. Data* **2,** 150–152 (2014).
19. Marzese, D. M. *et al.* Brain metastasis is predetermined in early stages of cutaneous melanoma by CD44v6 expression through epigenetic regulation of the spliceosome. *Pigment Cell Melanoma Res* **28,** 82–93 (2015).
20. Marzese, D. M. *et al.* DNA methylation and gene deletion analysis of brain metastases in melanoma patients identifies mutually exclusive molecular alterations. *Neuro Oncol* **16,** 1499–1509 (2014).
21. Orozco, J. I. J. *et al.* Epigenetic Profiling for the Molecular Classification of Metastatic Brain Tumors. *Nat. Commun.* https://doi.org/10.1038/s41467-018-06715-y (2018).
22. Hammond, M. E. *et al.* American Society of Clinical Oncology/College Of American Pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer. *J. Clin. Oncol.* **28,** 2784–2795 (2010).
23. Wolff, A. C. *et al.* Recommendations for human epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical Oncology/College of American Pathologists clinical practice guideline update. *J. Clin. Oncol.* **31,** 3997–4013 (2013).
24. Forbes, S. A. *et al.* COSMIC: somatic cancer genetics at high-resolution. *Nucleic Acids Res* **45,** D777–D783 (2017).
25. 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).
26. Triche, T. J. Jr., Weisenberger, D. J., Van Den Berg, D., Laird, P. W. & Siegmund, K. D. Low-level processing of Illumina Infinium DNA Methylation BeadArrays. *Nucleic Acids Res.* **41,** e90 (2013).
27. van der Maaten, L. J. P. & Hinton, G. Visualizing Data using t-SNE. *J. Mach. Learn. Res.* **9,** 2579–2605 (2008).
28. van der Maaten, L. Accelerating t-SNE using Tree-Based Algorithms. *J. Mach. Learn. Res.* **15,** 3221–3245 (2014).
29. Bengtsson, H. matrixStats: Functions that Apply to Rows and Columns of Matrices (and to Vectors). *R package version 0.53.1.* https://cran.r-project.org/web/packages/matrixStats/ (2018).
30. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis.* Springer-Verlag: New York (2009).
31. Neuwirth, E. RColorBrewer: ColorBrewer Palettes. *R package version 1.1-2* https://cran.r-project.org/web/packages/RColorBrewer/index.html (2014).
32. Krijthe, J. H. Rtsne: T-Distributed Stochastic Neighbor Embedding using a Barnes-Hut Implementation. *Github* https://github.com/jkrijthe/Rtsne (2015).
33. Morris, T. J *et al.* ChAMP: 450k Chip Analysis Methylation Pipeline. *Bioinformatics* **30,** 428–430 (2014).

Data Citations

1. Salomon, M. P., Marzese, D. M., Hoon, D. S. & Orozco, J. I. *Gene Expression Omnibus* GSE108576 (2017).
2. Hoon, D. & Huang, S. *Gene Expression Omnibus* GSE44661 (2013).

Acknowledgements

Author Contributions

M.P.S., J.I.J.O., D.S.B.H., and D.M.M conceptualized the study. M.P.S. processed DNA methylation data, performed data quality controls, and compiled and transferred data to NCBI's GEO. J.I.J.O. and A.O.M.-P. processed tissues and performed quality controls. J.S.W., P.H., C.S.C., and R.A.S. selected the patients, contributed with the tissue logistics, and annotation of clinical data. M.P.S. and D.M.M. constructed the figures. J.I.J.O. constructed the tables. D.S.B.H. and D.M.M. provided general guidance and oversaw the study. M.P.S., J.I.J.O., A.O.M.-P., D.S.B.H., and D.M.M. wrote the manuscript. All authors read, edited, and approved the final manuscript before submission.

Additional Information

Table 2 is only available in the online version of this paper.

Competing Interests: The authors declare no competing interests.

How to cite this article: Salomon, M. P. *et al.* Brain metastasis DNA methylomes, a novel resource for the identification of biological and clinical features. *Sci. Data*. 5:180245 doi: 10.1038/sdata.2018.245 (2018).

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.