Pike *et al.*
Acta Neuropathologica Communications (2024) 12:170
https://doi.org/10.1186/s40478-024-01874-0

Acta Neuropathologica Communications

RESEARCH

Open Access

Glioma immune microenvironment composition calculator (*GIMiCC*): a method of estimating the proportions of eighteen cell types from DNA methylation microarray data

Steven C. Pike[1,2,3], John K. Wiencke[4,5,6], Ze Zhang[2], Annette M. Molinaro[4,5], Helen M. Hansen[4], Devin C. Koestler[7], Brock C. Christensen[2,8,9], Karl T. Kelsey[10] and Lucas A. Salas[1,2*]

Abstract

A scalable platform for cell typing in the glioma microenvironment can improve tumor subtyping and immune landscape detection as successful immunotherapy strategies continue to be sought and evaluated. DNA methylation (DNAm) biomarkers for molecular classification of tumor subtypes have been developed for clinical use. However, tools that predict the cellular landscape of the tumor are not well-defined or readily available. We developed the Glioma Immune Microenvironment Composition Calculator (*GIMiCC*), an approach for deconvolution of cell types in gliomas using DNAm data. Using data from 17 isolated cell types, we describe the derivation of the deconvolution libraries in the biological context of selected genomic regions and validate deconvolution results using independent datasets. We utilize *GIMiCC* to illustrate that DNAm-based estimates of immune composition are clinically relevant and scalable for potential clinical implementation. In addition, we utilize *GIMiCC* to identify composition-independent DNAm alterations that are associated with high immune infiltration. Our future work aims to optimize *GIMiCC* and advance the clinical evaluation of glioma.

Keywords Glioma, DNA methylation, Tumor microenvironment, Epidemiology, Deconvolution

*Correspondence:
Lucas A. Salas
lucas.a.salas@dartmouth.edu
[1] Integrative Neuroscience at Dartmouth, Guarini School of Graduate and Advanced Studies at Dartmouth College, Hanover, NH, USA
[2] Department of Epidemiology, Geisel School of Medicine at Dartmouth College, Lebanon, NH, USA
[3] Department of Neurology, Dartmouth Hitchcock Medical Center, Lebanon, NH, USA
[4] Department of Neurological Surgery, University of California San Francisco, San Francisco, CA, USA
[5] Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, CA, USA
[6] Institute for Human Genetics, University of California San Francisco, San Francisco, CA, USA
[7] Department of Biostatistics & Data Science, Medical Center, University of Kansas, Kansas City, KS, USA
[8] Department of Molecular and Systems Biology, Geisel School of Medicine, Dartmouth College, Lebanon, NH, USA
[9] Department of Community and Family Medicine, Geisel School of Medicine, Dartmouth College, Lebanon, NH, USA
[10] Departments of Epidemiology and Pathology and Laboratory Medicine, Brown University, Providence, RI, USA

Introduction

Adult diffuse glioma represents the most common primary malignancy within the central nervous system (CNS), affecting an estimated 16,000 Americans each year [1, 2]. Pathological and molecular markers have historically classified this class of CNS tumors. Pathologists now rely on criteria set by the World Health Organization CNS tumor classification (WHO CNS5) to make diagnoses [3–5]. Isocitrate dehydrogenase (*IDH*) mutations were early adopted classification markers and strong markers of survival; lack of *IDH* mutation is associated with a dismal prognosis with a median survival of 14.6 months [1]. As of the most recent WHO CNS5 criteria, *IDH* mutations are used in combination with other findings to diagnose adult diffuse gliomas differentially and identify optimal treatment approaches [4, 5].

The immune microenvironment is a critical component contributing to tumor growth and survival, as well as an avenue for the treatment of many cancers. The tumor microenvironment in adult diffuse glioma consists of CNS resident populations (neurons, oligodendrocytes, astrocytes, and microglia), infiltrating immune cells, tumor vasculature, and structural support cells. The immune compartment consists of immune cells attempting to eliminate cancer cells or necrotic tissue (tumor-specific and non-tumor-specific T-cells, "proinflammatory" macrophages/microglia, neutrophils, etc.) or cells that the tumor is "hijacking" to promote tumor survival (regulatory T cells, myeloid-derived suppressor cells, "anti-inflammatory" macrophages/microglia, etc.) [6, 7].

There are predominant immunosuppressive factors in adult diffuse gliomas, particularly GBM, that impact the efficacy of immunotherapies [8]. This is driven by various factors, including the altered expression of cytokines and growth factors [9] and the upregulation of immunosuppressive myeloid populations [10]. These cells can differentially regulate cytokine and chemokine expression to recruit regulatory T cells, inhibit anti-tumor T cells, and stimulate exhaustion states in T cells [11]. The limited activity or presence of these T cells decreases the effectiveness of therapies that try to bolster antigen-specific responses, such as tumor vaccines or adoptive cell therapy [12, 13]. Therapies that try to evade the tumor immune suppression, such as immune checkpoint inhibitors, are receiving mixed results in clinical trials [14]. Although various therapies have initial successes, there is currently no FDA-approved immunotherapy for treating glioma.

There are many hypotheses as to why immunotherapy is not working for adult diffuse gliomas. Firstly, the anatomy of the CNS lends itself to unique immunological phenomena that don't exist for other tumor types, such as the blood–brain barrier, meningeal inflammatory pathways, and glymphatic system [15–17]. Secondly, many tumors have a low total mutational burden (TMB), which limits the pool of tumor antigens that an immune response can be enacted upon [18]. Lastly, there is significant variability in the tumor immune microenvironment [19]. Not only across subtypes of glioma but across tumors with the same diagnoses, we observe strong heterogeneity in the amount of vascularity, lymphocyte infiltration, and macrophage activity [20–22]. These underlying differences of the immune system, or how drugs can enter the tumor, may vary patient to patient. Understanding more about why immunotherapies fail or succeed on an individual patient level may allow us to conceptualize further how to optimize immunotherapeutic approaches [23, 24].

Currently, the predominant techniques to characterize the immune microenvironment cells include flow/mass cytometry, histopathology, or single-cell sequencing. Many of these options are challenging due to requirements for fresh samples, expensive equipment/reagents, optimized protocols, and staff with expertise. One technique that is being rapidly adopted in clinical practice, and can be utilized for immune profiling, is DNA methylation (DNAm) assays. DNAm refers to the epigenetic phenomena in which nucleotide bases are covalently modified by adding a methyl group; the most common loci for this are cytosine-guanine dinucleotides (CpGs) [25]. DNAm patterns in regulatory elements of genes can promote or inhibit the expression of genes. This mechanism can also lead to the activation of oncogenes or suppression of tumor suppressor genes in cancers and other disorders [25]. In addition to facilitating tumor progression, DNAm alterations are utilized during development to guide progenitor cells into terminally differentiated phenotypes [25–28]. Changes to DNAm allow for cell-type specific genes to be expressed, allowing for the diversity of all human cell types to be derived from the same genetic information. Researchers have identified patterns of cell-type specific DNAm and use this information to deconvolve bulk DNAm samples; in other words, to estimate the relative proportions of each cell type in the sample [29–34].

Utilizing DNAm data for scalable CNS tumor immune profiling is highly compatible with current developments in using this data for tumor classification. Various iterations of classifiers have been developed as this is a rapid area of research [35–38], including in 2018, where Capper et al. utilized patterns of DNAm microarray data to establish a neuropathology classifier that could predict tumor subtype [39]. This work shows the vast heterogeneity in epigenetic alterations in distinct

glioma subtypes but has strong clinical utility and can stratify patient populations. Therefore, building a high-throughput, cost-effective platform that can predict the immune environment of each sample in parallel to the tumor subtype would further leverage the already existing clinical utility of DNAm profiling and promote further research into optimizing both aspects in unison.

Currently, there are limited deconvolution libraries for the deconvolution of glioma DNAm data. Carcinoma deconvolution libraries have been introduced, such as *HiTIMED* [30], *MethylResolver* [40], and *methylCIBERSORT* [41, 42]; however, these methods do not include CNS-specific cell types. Additionally, single-cell DNAm data has been collected on CNS tumors; however, the feature space of this data is not compatible with microarray data. Our goal is to combine the best elements of all the previous methods with our recently validated brain deconvolution tool, *HiBED* [32], to develop DNAm deconvolution libraries for glioma that are subtype-specific and maintain a high resolution of both CNS and immune cells.

Materials and methods

Isolated cell type and glioma datasets

Detailed information about the source and donor demographics of the cell type-specific references are listed in Supplemental Table 1. In brief, this included the following CNS reference samples: human primary astrocytes from the post-mortem sub-ventricular deep white matter (Astro, n = 6) [43], endothelial (Endo, n = 12) and stromal (Stromal, n = 14) cells from umbilical cord tissue [44], GABAergic neurons (GABA, n = 5) and, glutamatergic neurons (GLU, n = 5) from the post-mortem dorsolateral prefrontal cortex [45], microglial cells from the post-mortem medial frontal gyrus, superior temporal gyrus, subventricular zone and thalamus (Micro, n = 18) [46], and oligodendrocytes from the post-mortem Brodmann area 46 (Oligo, n = 20) [47]. We utilized purified cell types obtained via flow and magnetic sorting methods for the immune cells as described in our previous publication [29]. In brief, this included neutrophils (Neu), monocytes (Mono), B naïve cells (Bnv), B memory cells (Bmem), CD4 naïve cells (CD4nv), CD4 memory cells (CD4mem), T regulatory cells (Treg), CD8 naïve cells (CD8nv), CD8 memory cells (CD8mem), natural killer cells (NK) as well as artificial mixtures of immune cell types. The healthy and anonymous donors included 41 males and 15 females, with a mean age of 32.2 years (SD = 12.2), and multiple race/ethnicities and further re-classified into broad genetic ancestries, including African (Sub-Saharan), East-Asian, Indo-European, and multiple/admixed. Horvath methylation age was inferred using Enmix software for any samples in which age was not provided [48, 49]. For samples without chromosomal sex data, this was inferred using *SeSAMe* software [50].

To integrate all of the reference data to compatible formats, i.e., converting whole genome bisulfite sequencing (WGBS) data to Illumina 450 k or EPIC data, we utilized *methylLiftover* [51]. Subsequently, we performed beta-mixture quantile normalization (BMIQ) [52] via *ChAMP* [53]. Probes that are known to be cross-reactive, closely associated with single nucleotide polymorphisms, exist on the X or Y chromosome, were on non-CpG sites, or had very large detection p-values [54] ($p > 0.01$) were removed from the analysis, resulting in a dataset of 126 samples across 306,466 CpGs.

This study utilized two large databases of DNAm data on glioma samples. The first dataset originated from the construction of glioma subtype-specific DNAm-based classifiers derived by Capper et al. [39]. Detailed information about the samples utilized from this dataset is available in Supplemental Table 2. In brief, this consisted of healthy brain samples across seven brain regions (n = 72) as well as glioblastoma (GBM; n = 671), grade 2/3 astrocytoma (AST; n = 172), grade 4 astrocytoma (AST-HG; n = 87), and oligodendroglioma (OLG; n = 163). The second dataset consisted of data produced by The Cancer Genome Atlas (TCGA), which has data on over 33 human cancer types that are publicly available (https://portal.gdc.cancer.gov/). Detailed information about the samples utilized from this dataset is available in Supplemental Table 5. In brief, this consisted of GBM (n = 132), AST (n = 139), OLG (n = 169) tumor samples alongside a set of samples that could not be confidently assigned to one of these groups (n = 216). These datasets were processed using *minfi* [55] via normal-exponential out-of-band background correction with dye bias normalization [56]. Probe filtering was done similarly to the cell-type-specific references.

Tumor diagnosis harmonization to WHO 2021 criteria

Since the datasets in this study were collected before the WHO 2021 revised classification of adult glioma implementation, we harmonized their previous annotations as such. For the Capper et al. dataset, we relied on the molecular classification of the tumors, which comprised four groups: glioblastoma, oligodendroglioma, astrocytoma IDH-mutant and astrocytoma IDH-mutant high grade (Supplemental Table 2). For the TCGA dataset, we relied on the availability of critical molecular markers to infer a WHO 2021 diagnosis in the legacy datasets (Supplemental Table 5). The TCGA offers glioma DNA methylation data under two large databases, one

labeled "GBM" for glioblastoma and one labeled "LGG" for low-grade glioma. Samples from the GBM dataset that were confirmed *IDH* wildtype were considered GBMs (n = 132). OLG samples were categorized by the presence of *IDH* mutation and 1p/19q co-deletion in the LGG dataset (n = 169). AST samples were identified as *IDH* mutant tumors in the LGG dataset that were not 1p/19q co-deleted. These samples were also negative for *TERT* promoter mutations, *EGFR* amplification, and 7 + /10- mutations (n = 139). The remaining samples were unable to be confidently mapped to one of our groups (n = 216).

GIMiCC library construction

L0 was constructed using the *InfiniumPurify* software [57] to identify the 1,000 most informative CpGs for discriminating healthy control brain samples from the glioma tumor samples from the Capper et al. dataset. The dataset was split 75:25 into training and testing; L0 was constructed using only the training set. Four libraries were produced for four subtypes of glioma according to Capper et al. defined molecular subtypes of adult diffuse glioma: GBM, OLG, AST, and AST-high grade (AST-HG). We developed all subsequent layers using an adapted version of the *meffil.cell.type.specific.methylation* function from the *meffil* package [58]. We selected 25 hybrid (hypo- and hyper-methylated) CpGs per cell type per library layer, as this was deemed most optimal in our previous studies [32].

GIMiCC hierarchy was designed to blend previous hierarchies developed for the human brain [32] and solid tumors [30]. L0 separates the tumor from the non-tumor fraction. Library 1 was built to deconvolve the non-tumor fraction into neuronal cells, glial cells, angiogenic cells, and immune cells. Furthermore, Layers 2A, 2B, and 2C were used to separate these broader cell categories into cell subtypes such as GABA and GLU from the neuronal cells, Astro, Oligo, and Micro from the glial cells, and Endo and Stromal from the angiogenic cells. Layer 2D separates the immune cell fraction into myeloid and lymphoid compartments, and further libraries were used to investigate even more specific subpopulations. Layer 3A splits the myeloid cells into Mono and Neu, and Layer 3B splits the lymphoid cells into NK, B cells, and T cells. Layer 4 is used to separate T cells into CD4 or CD8-positive T cells. Lastly, Layers 5A, 5B, and 5C are used to separate the B cell, CD8T, and CD4T cell compartments into memory and naïve subtypes, including Treg for CD4T cells.

GIMiCC hierarchical deconvolution

Firstly, the tumor and nontumor proportions are estimated from the probability density distribution of the L0 CpGs using the *InfiniumPurify* pipeline [57]. The projections for Layer 1 were calculated using the constrained projection/quadratic programming (CP/QP) approach developed by Houseman et al. [59], the sample is deconvolved into neuronal, glial, angiogenic, and immune cell proportions. These proportions are weighted by the proportion of non-tumor cells derived in L0 to develop the final deconvolution for Layer 1 into the tumor, neuronal, glial, angiogenic, and immune cell proportions. This process is then iterated for the rest of the hierarchical structure; project the cell proportions of layer n and weigh the estimates by the proportion of the parent node in the n-1 layer. Deconvolving to the deepest layer results in the full 18-cell type deconvolution.

Biological enrichment analyses

To map each CpG to associated genes, we utilized the InfiniumMethylation BeadChips Annotation file [60]. The UCSC Genome Browser was used to investigate the CpG location relative to associated gene(s) [61].

Gene set enrichment analysis was performed with the *goMeth* function from *missMethyl* software [62]. The input to the software is a list of differentially methylated CpGs and a total query CpG list. The output is a set of gene ontology (GO) terms with corresponding p-values for the test of enrichment. FDR values, generated with the Benjamini and Hochberg procedure, were used to select significantly enriched GO terms (FDR < 0.05).

Enrichment for relation to CpG islands was done with independent logistic regression models to calculate the odds of a query set of CpGs being in open sea, north shelves, north shores, islands, south shores, or south shelves. Similarly, we tested for enrichment in gene region to calculate the odds of a query set of CpGs being in TSS1500, TSS200, 5'UTR, 3'UTR, 1st exon, or gene body regions.

Benchmarking *GIMiCC* with other DNAm deconvolution methods

methylCIBERSORT [41] and *methylResolver* [40] were implemented as described. The glioma signature matrix was utilized for *methylCIBERSORT*. The default blood signature matrix was utilized for *methylResolver*. These methods, alongside *GIMiCC*, were used to deconvolve a set of artificial mixtures of immune cells (GSE182379) [29]. The error was quantified as the difference in the true versus the predicted proportions. Cell types were aggregated into five categories to equate similar cell types across methods: granulocytes (Gran), NK, Mono, T cells, and B cells.

Epigenome-wide association studies (EWAS)

EWAS analysis was done using the *minfi* and *limma* software [55, 63]. We utilized linear regressions to identify differentially methylated CpGs (DMCs) between two groups. Cell-type adjusted EWAS analysis includes the proportions of specified cell types, mean scaled and centered, as covariates in the linear model. DNAm proportions (β values) were converted to M-values by computing the logit of the β values in base 2. An empirical Bayes method was used to normalize CpG-wise residual variance. The Benjamini–Hochberg False Discovery Rate-FDR procedure was used to adjust for multiple hypothesis testing. Prior to the analysis, we defined a DMC by having an effect size ($|\Delta\beta|$) larger than 0.3 and an FDR less than 0.05.

Survival analysis

Cox-proportional hazards models were used to quantify the effect of cellular composition on survival in the TCGA dataset. Univariable and multivariable models were used; age and sex were included in the multivariable model. Additionally, separate models were created for each tumor type. The proportional hazard assumption was tested in each model; in cases where the assumption is not upheld in the multivariable model (adjusting for age and sex), results were derived from the univariable model. To visualize the results, we split the dataset based on the median value for a given cell type and compared the survival outcomes in the subpopulation with values above the median or "hot" for a given cell type to the subpopulation with values below the median or "cold" for a given cell type.

Results

GIMiCC hierarchical tree development and library construction

Our goal with this work is to develop a DNAm-based deconvolution method that can resolve the glioma microenvironment as a way to retrieve this information at epidemiological scales. Recent advances in our group have allowed for the development of DNAm-based deconvolution libraries for whole blood (*FlowSorted.BloodExtended.EPIC*) [29], brain samples (*HiBED*) [32], and solid tumors (*HiTIMED*) [30]. The construction of these libraries required the collection of cell-type specific reference DNAm data from isolated CNS and immune cell types. Thus, we combined these datasets to develop cell-type-specific reference profiles for significant expected cell types in glioma (Supplemental Table 1). We aggregated these datasets to create a database of 132 samples for 17 cell types, including GABAergic (GABA) and glutamatergic (GLU) neurons, oligodendrocytes (Oligo), astrocytes (Ast), microglia (Micro), endothelial cells (Endo), stromal cells (Stromal), *neutrophils* (Neu), monocytes/macrophages (Mono), natural killer cells (NK), T regulatory cells (Treg), naïve B cells (Bnv), memory B cells (Bmem), naïve CD4 + T cells (CD4nv), naïve CD8 + T cells (CD8nv), memory CD4 + T cells (CD4mem), and memory CD8 + T cells (CD8mem). When aggregating these data, we noticed the age distribution for microglia samples was much higher than the other samples. Because of the known epigenetic alterations associated with age in microglia and other phagocytic immune cells [64, 65], we stratified the samples into those coming from younger individuals (< 75) and those from older individuals (> 75). Because there was significant epigenetic variation between these two groups of samples, we opted to utilize only microglial samples from younger individuals in the deconvolution library development (Supplemental Figure 1).

From our previous experience with developing deconvolution strategies for complex tissues, we implemented a hierarchical deconvolution scheme. Hierarchical deconvolution resolves major cell types in more shallow layers and takes the results of those layers to scale the output of deeper layers. This approach forces our algorithm to leverage shared lineage marks across cell types independently of the markers used to resolve unique subsets, allowing us to resolve more cell types than previously [30]. We combined the hierarchical tree for CNS cells from *HiBED* with the hierarchical tree for the tumor-immune microenvironment from *HiTIMED* to generate a 6-layered hierarchical tree consisting of 17 cell types and a tumor cell proportion (Fig. 1). Layer 0 (L0) will first deconvolve the sample into the tumor and non-tumor cell proportions. In Layer 1 (L1), the non-tumor cell fraction is deconvolved into four broad categories: neuronal, glial, angiogenic, and immune. Subsequent layers then deconvolve each one of these categories into more specific cell types.

To generate the deconvolution library for L0, we utilized the Capper et al. dataset [39], which consists of DNAm-based classes of glioma in addition to healthy control samples across seven brain regions (Supplemental Table 2). The dataset was split into training and testing datasets in a 3:1 ratio, and each L0 was derived only using the training set. We generated an L0 library specific to distinct tumor types according to the most recent WHO CNS5 criteria: glioblastoma (GBM), oligodendroglioma (OLG), *IDH* mutant astrocytoma (AST), and high-grade *IDH* mutant astrocytoma (AST-HG). Prototype versions of *GIMiCC* that included a pan-glioma L0 were outperformed by the glioma-subtype-specific approach (data not shown). Each library consists of the 1,000 most informative differentially methylated CpGs (DMCs) when comparing the tumor samples to the

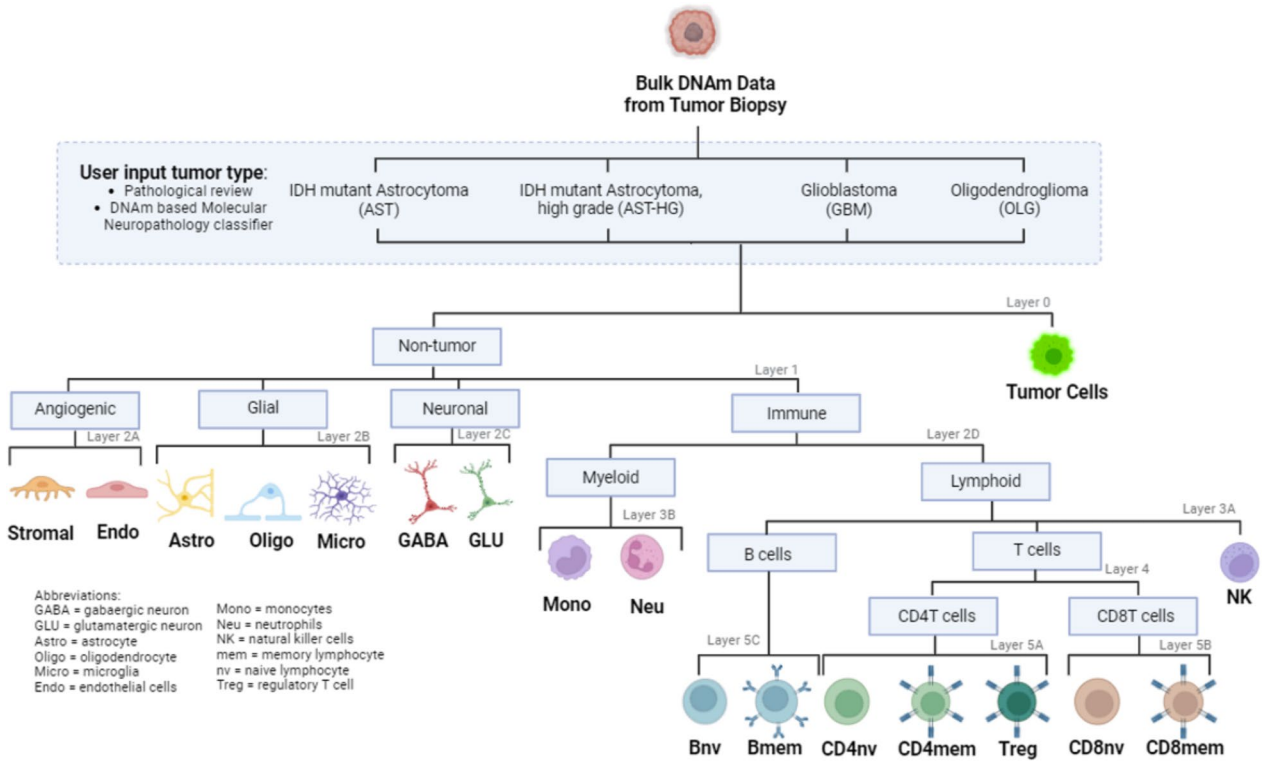


Fig. 1 Structure of *GIMiCC* hierarchical deconvolution using DNAm data from isolated cell types. The hierarchical structure allows users to specify the resolution in which they want to investigate the microenvironment, such as in a broad manner at Layer 2 into "myeloid" or "lymphoid" or in a specific manner at Layer 5, where details about B and T cell subtypes are distinguished. This figure was generated on biorender.com

healthy control (Fig. 2A). We found that the libraries for *IDH* mutant tumors were mainly hypermethylated in the tumor samples. Specifically, 99.7% of the OLG, 97.9% of the AST-HG, and 99.6% of the AST L0 CpG sites were hypermethylated in the tumors. In contrast, only 46.5% of the GBM L0 CpG sites were hypermethylated in tumors. In a subset of the Capper et al. dataset, the authors provided a tumor purity estimate using the Cancer Genome Atlas (TCGA) pan-glioma DNAm model [39, 66]. We found that *GIMiCC* tumor purity estimates correlate with the TCGA-based estimates (Fig. 2B–E). We additionally performed a sensitivity analysis in which we derived the AST L0 library using ten random splits of training and testing data for the tumors and controls. For the ten folds, 541 of the 1000 L0 CpGs were consistently identified in each iteration (Supplemental Figure 2A). Additionally, we show that the tumor purity estimates in the test samples were consistent with each iteration (Supplemental Figure 2B).

To generate the deconvolution libraries for all subsequent layers, we used the *limma* approach in the *Meffil* software [58] to select the top 25 hyper and hypomethylated cell-type-specific DMCs per cell type for each layer using linear models (Supplemental Figure 3–7). We observed minimal overlap between the CpGs utilized in each L0 library compared to the other deconvolution layers (Supplemental Figure 8).

The implementation of *GIMiCC* utilizes these derived libraries in a hierarchical fashion. Firstly, the tumor and nontumor proportions are estimated in the distribution of the L0 CpGs using the *InfiniumPurify* pipeline [57]. For all subsequent layers, the constrained projection quadratic programming approach [59] was used to project the proportions of the cell types in the subsequent layer by weighing their projections by the values of the previous projection. For instance, the outputs of using Layer 2A are weighted by the result of the neuronal cell proportion from using Layer 1. In this manner, we iterate this approach and estimate the proportions of all 18 cell types.

Biological context of CpGs selected in *GIMiCC* libraries

After developing the libraries for *GIMiCC,* we sought to understand which genome regions were utilized and connect these patterns to known biological functions. For instance, *VAMP2* is highly expressed in the brain as it is involved in synaptic vesicle fusion [67]. A CpG associated with *VAMP2* is included in the Layer 1 library

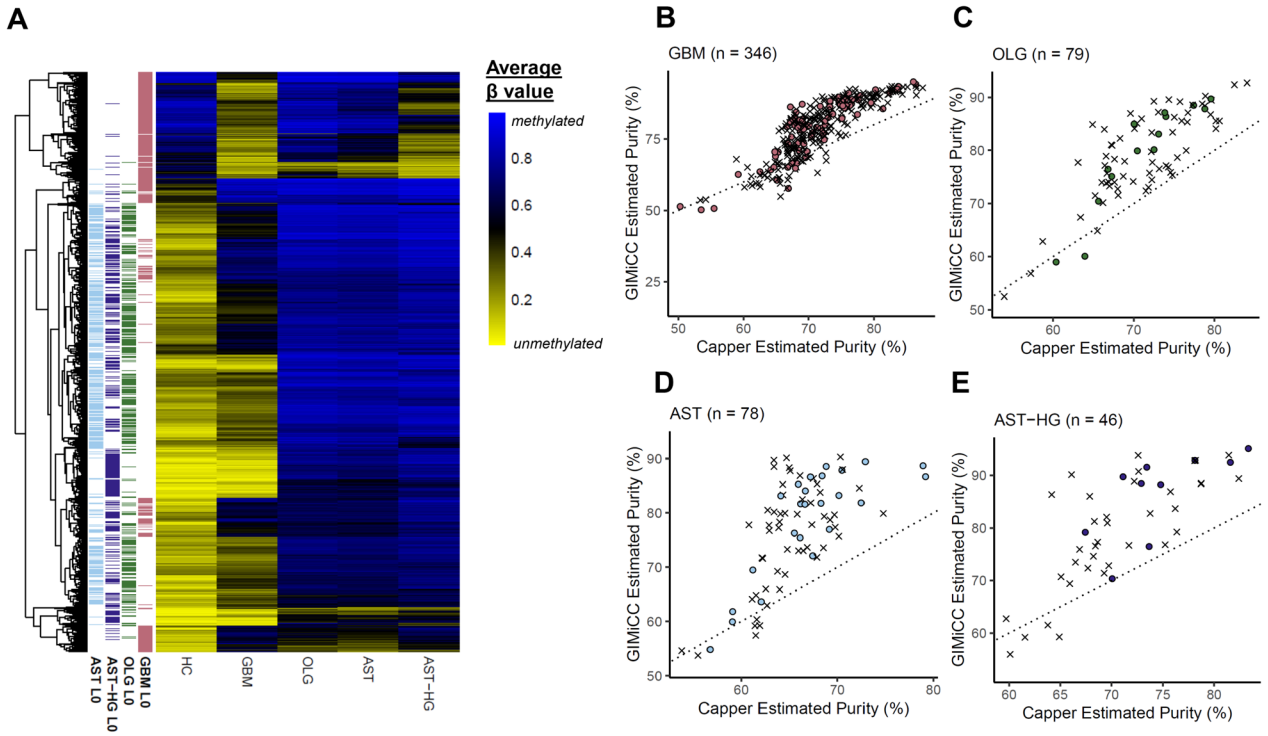


Fig. 2 CpGs selected for the L0 layer in the *GIMiCC* hierarchical deconvolution and accuracy of purity predictions. **A** Rows of the heatmap represent CpGs, the columns represent the average methylation value in the sample group in the test set of the Capper et al. dataset [39], and the color are representative of the methylation level of the sample at the specific CpG. The marks on the left of the figure denote which L0 library the CpG belonged to. Rows were organized via hierarchical clustering. **B–D** Correlation of *GIMiCC* estimated purity and the purity estimated in the Capper et al. manuscript [39, 66]. Circles denote samples from the test set, while crosses denote samples from the training set

and is hypomethylated in neurons compared to other cell types (Fig. 3A). Similarly, hypomethylation of a CpG near *TREM2* in microglia is used in Layer 2 of the deconvolution (Fig. 3B). *TREM2* is a marker of microglia and macrophages [68, 69]; however, on the epigenetic level at this site, it seems to be specific to microglia. Lastly, a CpG near *HLA-DOB*, a major histocompatibility complex gene highly expressed on B cells [70], is used in Layer 3 of the deconvolution (Fig. 3C).

We examined the compatibility of *GIMiCC* across the three most recent Illumina DNAm profiling arrays: 450k, EPIC, and EPICv2. We identified that most, but not all, CpGs were conserved across platforms (Supplemental Table 3). To determine if this would impact the results of *GIMiCC,* we compared the deconvolution results of the Capper et al. dataset with the entire probe set and with the probes only available on EPICv2. We found that these results were strongly correlated, with a median correlation coefficient of 0.99 and a minimum of 0.96 (Supplemental Figure 9).

To get a pathway-level perspective of the CpGs utilized in the *GIMiCC* libraries, we performed a gene set enrichment with the *missMethyl* software [62]. For all L0 libraries, the significantly enriched pathways and ontologies involved the plasma membrane and cell adhesion, whereas the GBM L0 library was also enriched with immune-related processes (Supplemental Table 4). For the libraries associated with the other layers of deconvolution, the enriched pathways and ontologies involved immunological pathways and processes (Supplemental Table 4).

Next, we tested for enrichment in the *GIMiCC* library CpGs for contexts of the CpG island methylation region (island, shore, shelf, and open sea) [71]. Open-sea CpGs were most enriched in the GBM L0 library, OR = 1.3, 95% CI [1.2–1.5] (Fig. 3D). However, CpGs within CpG islands were most enriched in the AST L0 library, OR = 3.7 [3.2–4.2], AST-HG L0 library, OR = 7.7 [6.6–8.9], and OLG L0 library, OR = 4.1 [3.6–4.7], (Fig. 3E, Supplemental Fig. 10). The CpGs in cell-type specific layers (L1-L5) were enriched on open sea CpGs, OR = 1.5 [1.4–1.8], and were unlikely to be on CpG Islands, OR = 0.59 [0.51–0.69], (Fig. 3F).

Lastly, we tested for enrichment of *GIMiCC* library CpGs within gene regulatory regions (TSS1500, TSS200, 5'UTR, 1st exon, Body, 3'UTR). We found that all L0

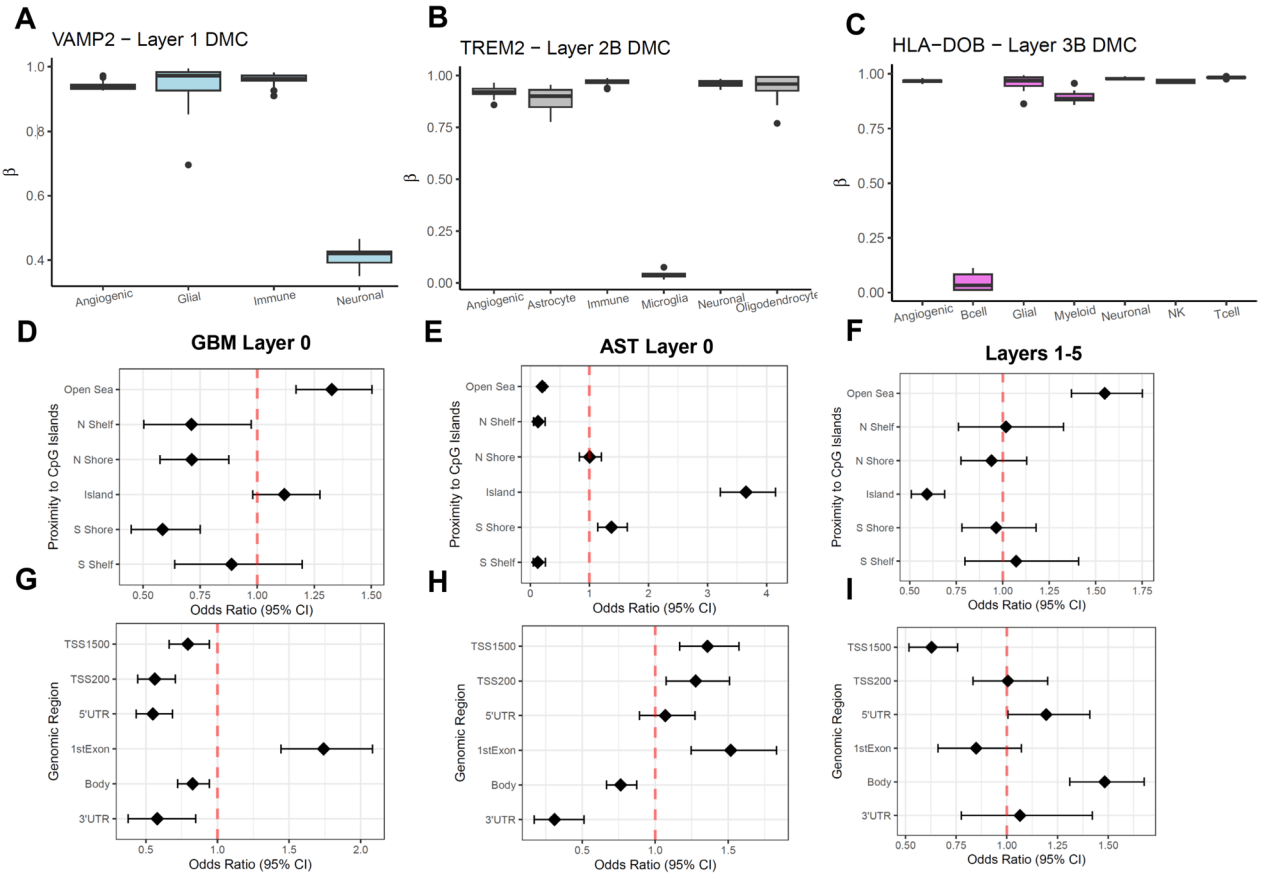


Fig. 3 Genomic context of CpGs utilized in *GIMiCC*. **A–C** The methylation level of representative CpGs from *GIMiCC* libraries at functionally relevant genes. **D–F** Enrichment analysis testing for the odds of the DMCs of the GBM L0 (**D**), AST L0 (**E**), and all other deconvolution layers (**F**) being in a CpG island, shore, shelf, or open sea region. **G–I** Enrichment analysis testing for the odds of the DMCs of the GBM L0 (**G**), AST L0 (**H**), and all other deconvolution layers (**I**) being within a particular gene regulatory element. *N* north, *S* south, *UTR* untranslated region, *TSS* transcriptional start site

libraries were enriched with CpGs on the 1st exon of genes; GBM OR = 1.7 [1.4–2.1], AST OR = 1.5 [1.2–1.8], AST-HG OR = 1.7 [1.4–2.1], and OLG OR = 1.4 [1.2–1.7], (Fig. 3G, H, Supplemental Fig. 10). There was enrichment for CpGs within gene bodies in the remaining L1-L5 layers OR = 1.5 [1.3–1.7] (Fig. 3I).

GIMiCC tumor type specificity

Using the test subset of samples, we projected the tumor purity using all four L0 libraries and compared these estimates (Fig. 4A). We found that the AST-HG, AST, and OLG L0 libraries yielded similar estimates. Additionally, using an AST-HG, AST, or OLG L0 to deconvolve GBM tumors results in lower tumor purity estimates. Similarly, the distribution of purity estimates is much lower using a GBM L0 to deconvolve OLG or AST tumors. Interestingly, AST-HG tumors deconvolved with any of the four L0 libraries yield similar estimates.

Qualitative and semi-quantitative validation of *GIMiCC*

We observed low tumor purity when *GIMiCC* was applied to non-tumor samples (Fig. 4A). Additionally, when we look at the projections of the other cell types, we observe cell proportion estimates that reflect the physiology of the tissue type, i.e., a high proportion of oligodendrocytes in the white matter (WM) samples, the highest levels of neurons in cortical samples (HEMI), and high levels of angiogenic and immune cells in more vascularized regions of the brain such as the anterior pituitary gland (ADENOPIT), pineal gland (PINEAL) and the pons (PONS).

We wanted to further validate the ability of *GIMiCC* to detect immune infiltration in brain tissue. To do this, we utilized additional samples from the Capper et al. dataset that were annotated to have "high immune infiltrate" and "low-yield" (Fig. 4C, D, Supplemental Table 2). For the former, these samples have a high granulocytic infiltration associated with necrosis or intense

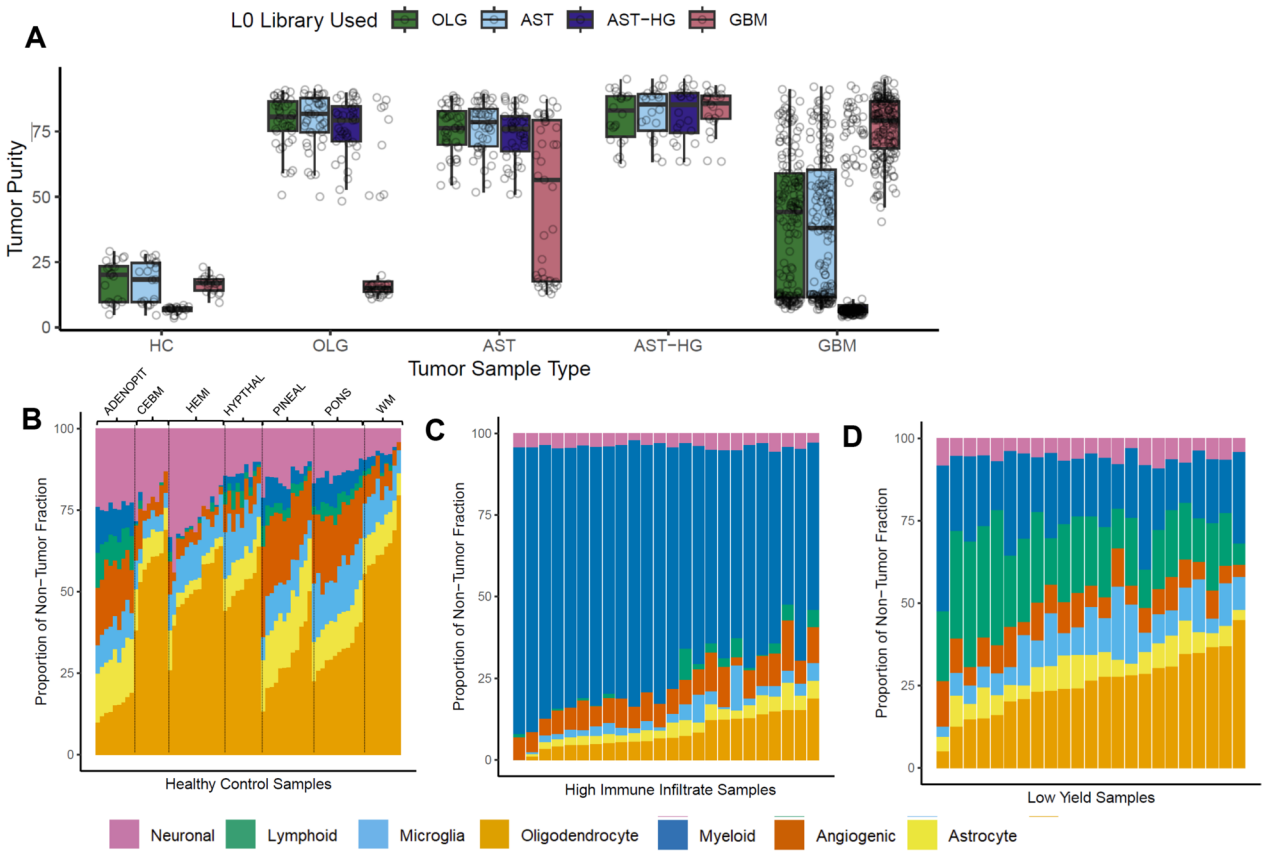


Fig. 4 *GIMiCC* predictions in Capper et al. test set and healthy controls. **A** Boxplots show the distribution of the sample's predicted cell fractions, tumor cells. The x-axis represents the group of samples, whereas the boxplot color represents the different L0 library used to derive the purity **B** *GIMiCC* predictions of the composition of the healthy control samples in the Capper et al. dataset. Each column represents a single sample. The y-axis represents the scaled proportion of the non-tumor portion. *ADENOPIT* anterior pituitary gland, *CEBM* cerebellum, *HEMI* cortical hemisphere, *HYPTHAL* hypothalamus, *PINEAL* pineal gland, *PONS* pons, *WM* white matter. *GIMiCC* predictions of the composition of **C** the high immune infiltrate samples and **D** low-yield samples in the Capper et al. dataset

hemorrhage. The latter is comprised of samples that had very low tumor cell content. We show that *GIMiCC* correctly predicts high fractions of immune cells within these samples compared to the healthy control samples.

To validate *GIMiCC* with an independent dataset, we used the DNA methylation data from the Cancer Genome Atlas (TCGA, https://www.cancer.gov/tcga). We selected genetically confirmed tumors and harmonized their diagnoses to WHO 2021 criteria (Supplemental Table 5, See Methods). Using *GIMiCC*, we estimated these samples' cellular composition, including the immune microenvironment (Fig. 5, Supplemental Figs. 11–13, Supplemental Table 6). The OLG tumors had much lower immune infiltration than the other tumor types, and the GBM tumors had higher levels of infiltration. The predominant infiltrates were myeloid, including microglia, neutrophils, and monocytes.

To further validate *GIMiCC*, we compared our estimates of tumor purity to the TCGA a consensus purity estimate (CPE); this estimate integrates predictions based on gene expression data, somatic copy-number data, DNAm data, and immunohistochemistry to generate a single estimate for tumor purity [72]. Our tumor purity estimates were highly correlated with the TCGA CPE (Fig. 6A–C).

Lastly, we compared the immune composition output of *GIMiCC* to the ESTIMATE immune score, which uses expression data to identify the level of immune cells in the TCGA tumors [73]. The samples deemed highly inflamed by *GIMiCC* have an elevated ESTIMATE immune score compared to the others (Fig. 6D). This relationship is emphasized further when incorporating microglia into the immune composition (Fig. 6E).

The ideal experiment to test the performance of DNAm-based deconvolution methods is to use a validated sorting method such as flow cytometry to

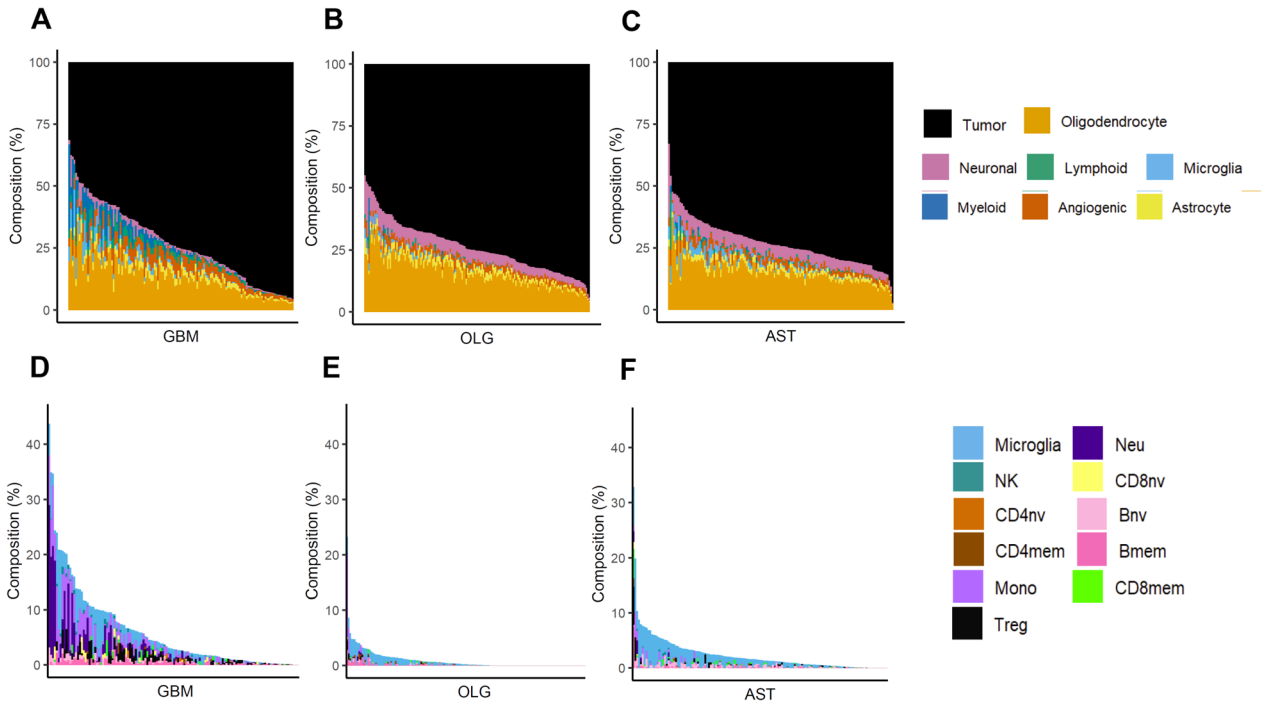


Fig. 5 *GIMiCC* predictions in the TCGA glioma samples. Each column represents a single sample, and the y-axis represents the proportion of cells within each tumor sample. **A–C** Lower layer deconvolution results for four subtypes of brain tumors. **D–F** Deeper layer deconvolution of the immune microenvironment of each tumor type

isolate each cell of interest and to develop artificial mixtures of known proportions to deconvolve them computationally. To our knowledge, no dataset as such is yet available. Thus, we utilized a dataset of artificial DNA mixtures of immune cells from human blood to test the ability of *GIMiCC* to deconvolve the immune microenvironment. We compared *GIMiCC* to two validated deconvolution methods: *methylCIBERSORT* [41] and *methylResolver* [40]. We found that *GIMiCC* performed at the same level as these other methods (Supplemental Fig. 14).

Determining the impact of cellular heterogeneity on DNAm-based tumor classification

Previous work has shown that glioma subtypes can be identified via distinct DNAm patterns; however, we hypothesized that these classifiers might be confounded by differences in cellular composition across tumor types when constructing tumor classifiers [37]. To test this, we performed epigenome-wide association studies (EWAS) comparing different tumor types from the Capper et al. dataset to generate lists of differentially methylated CpGs (DMCs) that can be used for tumor classification. To identify the effects of controlling for cellular heterogeneity, we ran five different EWAS models: 1) unadjusted for cell type, 2) adjusted only for the tumor fraction, 3) adjusted for broad cell categories {Tumor, Angiogenic, Glial, Immune} 4) adjusted for the immune microenvironment {Tumor, Angiogenic, Astrocyte, Microglia, Oligodendrocyte, Myeloid, Lymphoid} and 5) adjusted for a deeper layer deconvolution of the immune microenvironment {Tumor, Angiogenic, Astrocyte, Microglia, Oligodendrocyte, Tcell, Bcell, NK, Neu, Mono}. We identified significant heterogeneity in cellular composition within tumor types in the Capper et al. dataset (Supplemental Fig. 15, Supplemental Table 7). In an EWAS comparing GBM to AST gliomas, we identified more than 30,000 DMCs in the unadjusted model; however, there was no significant change to the number of identified DMCs when adjusting for cell type (Fig. 7A–C). We repeated this analysis by comparing other types of tumors to each other and found similar results (Supplemental Fig. 16). We found that identifying CpGs for tumor classification was not improved via controlling for cellular composition.

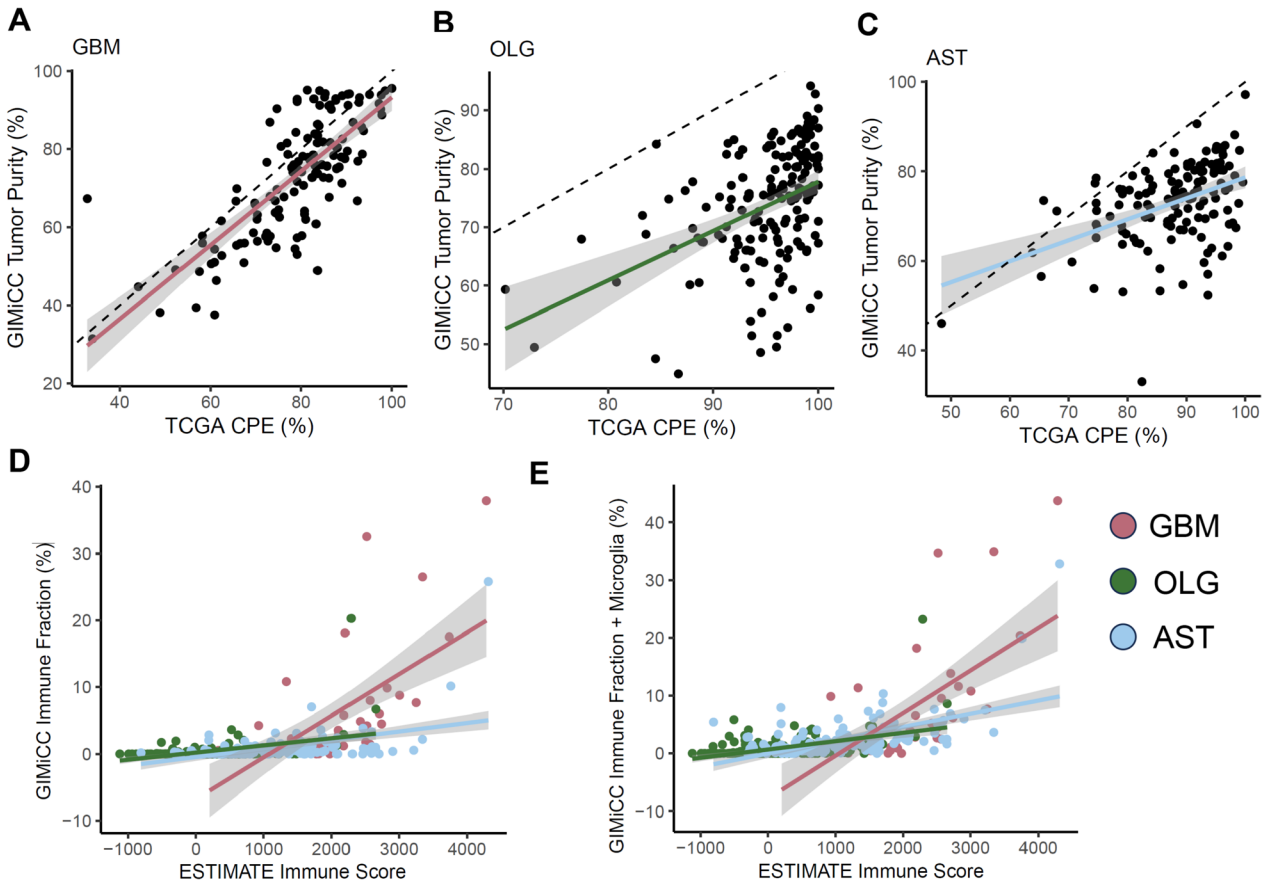


Fig. 6 Semi-quantitative validation of *GIMiCC* predictions in the TCGA dataset. **A–C** Correlation of consensus purity estimate (CPE) [72] with the *GIMiCC* predicted tumor proportion. **D** The relationship between ESTIMATE immune scores [73] and by *GIMiCC* derived immune estimates. Colors represent the three tumor types (red = GBM, green = OLG, blue = AST) **E** Visualizing the same relationship, but including microglia into the microenvironment

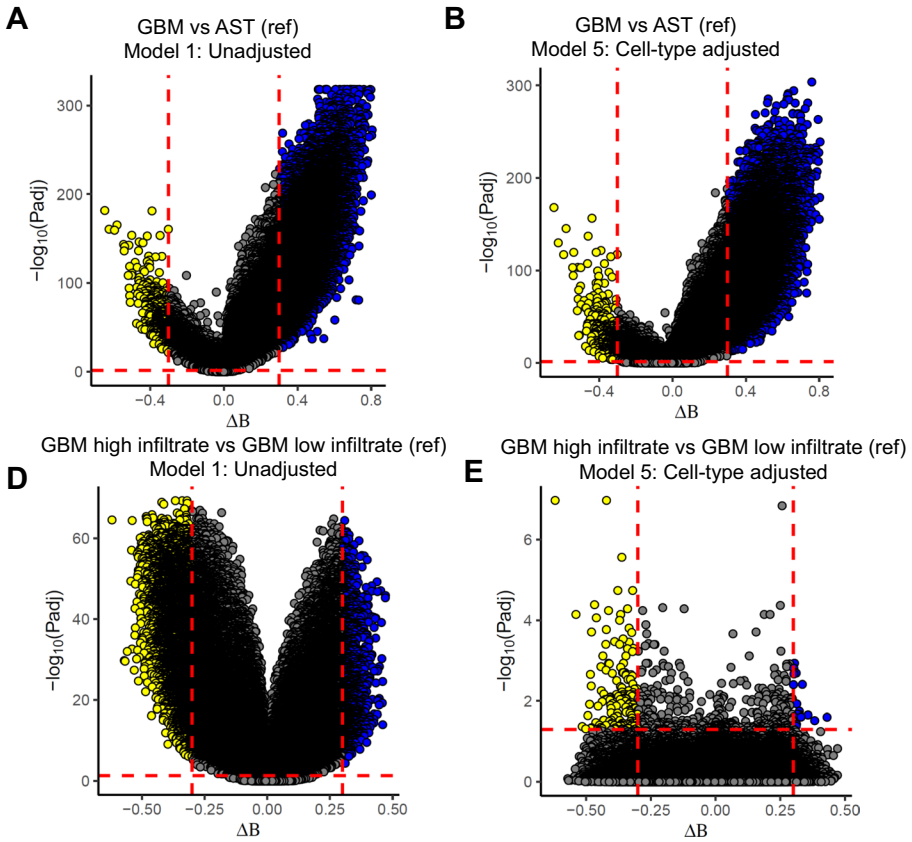
Identifying compositionally independent alterations in DNAm associated with highly infiltrative glioma

Because we identified a high variation of immune infiltration within tumor types, we aimed to identify DNAm patterns associated with increased inflammation in the Capper et al. GBM tumor samples, we performed an EWAS analysis comparing the highest and lowest deciles of immune infiltrated samples. Using the five models described previously, we show that controlling for cell type in this analysis dramatically reduces the number of DMCs identified (Fig. 7D–F). We also observed the same trend in the other tumor types (Supplemental Fig. 17).

Assessing the clinical relevance of *GIMiCC* estimates

Lastly, we sought to understand whether *GIMiCC*-derived immune cell proportions were associated with patient survival. To do this, we built Cox proportional hazard models to test the survival effects in the TCGA samples. We constructed univariable and multivariable models for each cell type to test the relationship between immune cell level and patient survival after 5 and 10 years (Supplemental Fig. 18–22). The multivariable model adjusted for age and sex. The models were also stratified by tumor type as including tumor type in a global model produced large violations of proportional hazards assumptions (data not shown).

This analysis showed that higher levels of angiogenic cells were most strongly associated with worse survival outcomes in OLG and AST tumors (Fig. 8, Supplemental Fig. 18–22). For GBM, we observed that the presence of many immune cell types, including CD8nv, Mono, and Microglia, was associated with worse survival. Interestingly, microglia are associated with survival but only in the multivariable model adjusting for age and sex ($HR_{uni} = 1.06$ [1.00–1.13] compared to $HR_{multi} = 1.16$ [1.08–1.25]).



C

	# DMCs	
	Hyper	Hypo
Model 1	31,059	456
Model 2	31,059	456
Model 3	31,059	456
Model 4	31,059	456
Model 5	31,059	456

F

	# DMCs	
	Hyper	Hypo
Model 1	1,231	7,413
Model 2	280	1,670
Model 3	62	507
Model 4	58	318
Model 5	20	154

Fig. 7 DMCs independent of cellular composition associated with high tumor immune infiltration. A–C EWAS analysis comparing GBM glioma to AST glioma in the Capper et al. dataset using **A** Model 1 and **B** Model 5. **C** Summarized results of all models used to compare tumor types. **D–F** EWAS analysis comparing the highest immune infiltrating GBM glioma to the lowest immune infiltrating GBM glioma in the Capper et al. dataset using **D** Model 1 and **E** Model 5. **F** Summarized results of all models used to compare high and low infiltrating GBM gliomas. Model 1: Univariable analysis. Model 2: Controlled for tumor proportion. Model 3: Controlled for proportions of tumor, angiogenic, glial and immune cells. Model 4: Controlled for proportions of tumor, angiogenic, astrocyte, microglia, oligodendrocyte, myeloid, and lymphoid cells. Model 5: Controlled for proportions of tumor, angiogenic, astrocyte, microglia, oligodendrocyte, T cell, B cell, NK, neutrophil, and monocyte cells. *Ref* reference group, *DMCs* differentially methylated CpGs, *Hyper* number of hypermethylated DMCs, *Hypo* number of hypomethylated DMCs

Discussion

Motivated by our group's recent successes in developing DNAm-based deconvolution libraries for whole blood [29], brain [32], and solid tumors [30], we aimed to integrate the cell references to have a set of deconvolution libraries for human adult diffuse gliomas. We identified cell-type-specific methylation patterns in 18 unique cell types and used these signatures to predict the composition of glioma samples hierarchically across six total layers. This new method was validated in an excluded subset of the training data (Capper et al.) and an independent dataset (TCGA) and compared to other DNAm-based deconvolution methods. We show that *GIMiCC* recapitulates the known biology of glioma inflammation, and its predictions can be used to learn more about the diversity of glioma immune microenvironments and perform cell-type adjusted EWAS analyses.

GIMiCC prediction of tumor purity

We developed four versions of L0 for subclasses of adult diffuse glioma defined by DNAm profile (AST, AST-HG, OLG, and GBM). We were motivated to create glioma subtype-specific deconvolution, as it produces more accurate results than a pan-cancer tumor purity estimation [30].

Mutations to *IDH* result in metabolic reprogramming of the cell, causing an accumulation of D-2-hydroxyglutarate (D-2-HG), which can disrupt the demethylation of histones and DNA [74]. This has been linked to a distinct hypermethylated phenotype, particularly in CpG islands [75, 76]. Our results show that the CpGs used in L0 for the *IDH* mutant gliomas were all hypermethylated, and our enrichment analysis showed high enrichment for L0 CpGs to be on CpG islands. This further validates our approach as the L0 libraries reproduce the known biology of these tumors.

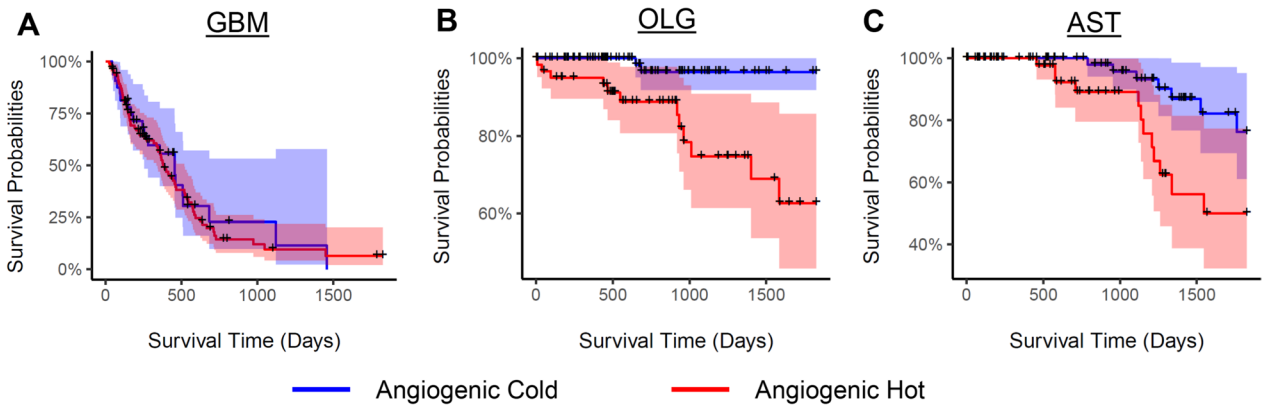


Fig. 8 Higher levels of angiogenic cells within the tumor microenvironment are associated with worse 5-year survival in TCGA samples. Kaplan–Meier curves are stratified by the median percentile of angiogenic cells where if a sample is above the median, it is assigned to the "hot" (red) strata while the remaining is assigned to the "cold" (blue) strata. Individual models were run for each tumor type: **A** GBM angiogenic hot (n = 96, median angiogenic proportion = 4.7) versus GBM angiogenic cold (n = 32, median angiogenic proportion = 1.5), **B** OLG angiogenic hot (n = 62, median angiogenic proportion = 3.3) versus OLG angiogenic cold (n = 104, median angiogenic proportion = 1.6) and **C** AST angiogenic hot (n = 57, median angiogenic proportion = 3.5) versus GBM angiogenic cold (n = 80, median angiogenic proportion = 1.5)

Using the test set samples from the Capper et al. dataset, we show that the L0 libraries predicted low tumor purity when applied to the incorrect tumor type. However, subsets of samples exhibited high purity regardless of the library utilized. This may reflect the nature of gliomas to exhibit a heterogeneity of tumor cell types that may not be fully captured by a single cell-type profile [77–79]. A fundamental limitation of our method is the need for more flexibility to account for mixtures of tumor cell types. Further studies utilizing single-cell DNA methylomics are required to extend *GIMiCC* and potentially deconvolve the tumor fraction.

We heavily relied on the resources of the TCGA database to validate *GIMiCC* L0 predictions. The CPE-predicted tumor purity is a well-established benchmark for tumor purity estimates in this database [72]. This method integrates tumor purity estimation via four platforms: RNA expression data [73], copy number alteration data [80], DNAm data [72], and immunohistochemistry [72]. Our results were highly correlated with this estimate, with the highest correlation for GBM L0. The *IDH* mutant L0 libraries tended to underpredict tumor purity compared to CPE.

GIMiCC prediction of immune composition

Layers 1 through 5 of *GIMiCC* are used to predict the remaining composition of the tumor samples. We could test the accuracy of this prediction with artificial mixtures of immune cells constructed independently. Compared to other methods of DNAm-based deconvolution, we concluded that *GIMiCC* performed at the same level, if not better, than the other approaches. Additionally, we showed that the performance of *GIMiCC* is stable using the probes available on the next iteration of the Illumina DNAm sequencing platform, EPICv2 [81].

Some of the CpGs included in *GIMiCC* are associated with genes expected to be functionally active in only particular cell types. Firstly, a CpG associated with *VAMP2* was included in Layer 1 to distinguish neurons from other cell types. *VAMP2* plays a role in vesicle fusion at the synapse during neurotransmission [67]; thus, we would expect to find the region around this gene hypomethylated in neurons. Secondly, we highlighted a CpG near *TREM2* that was included in Layer 2 to separate microglia from other cell types. *TREM2* is expressed on microglia and tissue-resident macrophages and is a significant target for current research about microglial immunology and neuroinflammation [82]. Although peripheral monocytes/macrophages express *TREM2* [83], this particular CpG is only hypomethylated in microglia. Our results suggest that *TREM2* may be differentially regulated epigenetically across peripheral (monocyte) versus local (microglia) cell types. Lastly, we identified a B cell-specific hypomethylation of a CpG near *HLA-DOB* known to be overexpressed on B cells [70]. These results, taken together with the gene ontology data, show that the libraries used in *GIMiCC* are connected to known biological processes of the cell types.

We additionally validated *GIMiCC* using it on inflammatory and low-yield control samples from Capper et al. Verbatim descriptions of these groups can be found in Supplemental Table 2 and the original publication [39]. Both groups of samples were noted to be high in leukocytes, either due to necrosis, hemorrhage, or tumor infiltration. *GIMiCC* can detect

the immune signal in these samples compared to true healthy controls, further validating the method.

Our last form of qualitative validation comes from comparing our results to what is currently known about the immunological heterogeneity across brain tumors. Across methods including single-cell RNA-sequencing [20, 84–88], imaging mass spectroscopy [87], flow cytometry [89, 90], and mass cytometry [91], there is significant variation in the immune composition across gliomas, which we observe in both Capper et al. and TCGA datasets. Even within distinct tumor types, there is known heterogeneity [92, 93]. One potential source of this variation in these datasets is tumor progression, as it has been shown that the tumor microenvironment alters over time, increasing with signatures of oligodendrocytes and myeloid cells [88]. Another potential bias could be in which section of the tumor was used for DNAm array, as the tumor core and periphery have distinct immune microenvironments [87, 90].

Interestingly, we also have replicated known findings about the differences in immune microenvironments between tumor subtypes. For instance, glioblastoma samples were generally more infiltrative than the other tumor types, consisting primarily of infiltrating monocytes and T cells that likely contribute to local immunosuppression [20, 84, 87, 94]. Microglia was the largest immune microenvironment contributor in the *IDH* mutant tumors, whereas in the GBMs we observed high monocyte infiltration. This aligns with previous findings indicating bone marrow-derived macrophages are increased in higher grade and immunosuppressed tumors, but microglia are present at all grades [84, 90, 95]. A limitation of *GIMiCC* is the ambiguity of tumor-associated macrophages and whether these cells are classified as "microglia" or "monocytes." The ability of cells to preserve epigenetic markers of origin has previously been identified in our group [96, 97]. We hypothesize that because these cells derive from two distinct lineages (microglia from yolk-sac progenitors versus monocytes from myeloid progenitors in the bone marrow) [98], distinct epigenetic marks would exist that maintain information about the source progenitor. Thus, we expect that monocytes that infiltrate and differentiate into tumor-associated macrophages in the brain would be captured in the monocytes ("Mono") proportion of *GIMiCC*; however, this has yet to be shown.

When looking at the adaptive immune cells, we find higher levels of T and B cells in GBM tumors compared to the other subtypes, which have been replicated in other studies [20, 85]. In particular, we corroborate other's findings of higher levels of mature and regulatory T and B cells in GBMs [85, 91, 99, 100]. It is worth noting that these cell types were only detectable in a subpopulation of patients within each tumor type, which may explain the variability in the efficacy of certain immunotherapies.

Impact

Using *GIMiCC*, we were able to address two main hypotheses about the cellular heterogeneity of glioma and its impact on DNAm analyses.

Firstly, we explored whether the composition of the microenvironment could influence the DNAm-based classification of glioma. Others have shown that resolving the non-tumor fraction of the sample aids in the correct assignment of tumor class when using DNAm biomarkers [37]. We tested this by performing EWAS analyses between distinct tumor types that did and did not control the tumor's cellular heterogeneity. We found that adding this information to the model did not change the calling of DMCs. Thus, we conclude that for strong molecular alterations such as *IDH* mutations, DNAm-based classifiers can select CpGs that are less likely to be impacted by compositional variation. However, these findings may not be true when trying to distinguish less pathologically distinct tumor types with DNAm data.

Secondly, we wanted to identify specific DNAm alterations associated with higher immune infiltration. Similarly, we performed multiple EWAS analyses comparing high to low infiltrating tumors while adjusting or not adjusting for cell type. In this case, adjusting for cell heterogeneity greatly reduced the number of DMCs being called. Without controlling for cell proportions, many of the DMCs are associated with the proportion of a given immune cell rather than the distinct biological alteration that is causing the downstream recruitment of leukocytes [101]. By running a model that can fully adjust for this, we can more readily point to the molecular alterations associated with recruitment without the confounding of having more cells in the "high infiltrate" group.

Lastly, we utilize *GIMiCC* to identify levels of cell proportions that are associated with 5 and 10-year-long survival. Interestingly, we find that higher levels of stromal and/or endothelial cell proportions in the tumor are associated with worse survival in OLG and AST, but not GBM. Marks of angiogenesis and microvascular proliferation are included in the diagnostic criteria of GBM; thus, we suspect we did not observe this association due to lower variance in the presence of angiogenic cells (Fig. 5). Nonetheless, this finding has been replicated in independent modalities in the TCGA set of tumors [102, 103]. Our previous work in solid tumors has also shown a similar effect in head and neck squamous cell carcinoma, stomach

adenocarcinoma, and thyroid carcinoma; higher levels of angiogenic cells in these tumors were associated with worse 5-year survival [30]. Other work on glioma has replicated this finding [102], including the observation that increased *Angiogenin* (*ANG*) expression correlates with worse survival outcomes and more immune infiltration [104]. These results suggest that therapies targeted toward preventing angiogenesis within these tumors may improve survival by restricting nutrients or immunosuppressive cells from the tumor.

Limitations

We have identified several limitations to our study. The first is that cell-type specific references were aggregated from various sources across many different DNAm profiling platforms. Although we address this in our processing pipeline, this inherently increases the risk of batch effects.

Second, proper implementation of *GIMiCC* requires prior knowledge of tumor classification. We show that using the wrong library to deconvolve a glioma subtype may deflate the tumor purity estimation. Thus, any misclassification of tumors will greatly impact the interpretation of the results, which may also explain the presence of outliers in the datasets we assessed.

Third, *GIMiCC* was built using data where tumor type was annotated via DNAm profile as defined in the Capper et al. dataset. This limits our resolution in tumor grade being able to distinguish grade 2/3/4 astrocytoma individually. As additional annotations for DNAm data are developed, we will be able to rapidly incorporate novel, more accurate classification schemes into our pipeline.

Fourth, the cell types included in *GIMiCC* are built off pre-identified cell populations with established markers and isolation protocols. We are limited in the ability to detect novel cell populations until we can derive an isolated DNAm cell profile. As our knowledge of tumor immunology grows and we can develop profiles for more specific cell types (such as myeloid-derived suppressor cells), we can incorporate them into our algorithm to efficiently and systematically screen various large databases for these new cell types.

Lastly, we have validated *GIMiCC* using publicly available datasets with silver-standard purity estimates and artificial mixtures of immune cells. We expect that direct matching of DNAm data to other modalities for validation (single-cell RNA sequencing, chromatin accessibility, or histology) will bias our estimates of cell proportion for two reasons. One is that the microenvironment cells that get captured will vary across separate regions of the tumor or serial slices of the tumor [20]. The second reason is the lack of overlap in feature spaces between DNAm and these other modalities. Single-cell approaches do not guarantee the detection of marker genes across all cell types, which may bias cell type identification. When collecting single-cell information, technical and biological variations due to cell cycle and cell states may increase the chance of classical measurement errors, biasing the results to the null and reducing precision. In contrast, deconvolution approaches assume that the average of the weighted signal represents a layer of a cell hierarchy. This increases the power of the analysis but changes the estimates, introducing what is known as a Berkson error (increased variability but less systematic biases when used correctly) [105]. Our strongest form of validation of *GIMiCC* would require the development of artificial mixtures of DNA from each cell type across a robust number of patients, including neuronal, immune, and tumor cell types, like our approach in Supplementary Fig. 11. As more specific DNAm profiles become available in the literature, our platform is flexible for the addition of updated cell types to the library or calibration to a true validation set.

Summary

We have introduced *GIMiCC*, a computational tool that allows users to calculate the cellular composition of glioma samples from DNAm data. We have shown that this method is valid and benchmarked to previous DNAm deconvolution methods while increasing the output's resolution and developing glioma-subtype specificity. As DNAm data is being collected more frequently for tumor identification, we plan to continue to optimize *GIMiCC* and implement it to understand more about the immunological interactions in glioma and potentially add information regarding optimizing treatment strategies.

Abbreviations

GIMiCC	Glioma immune microenvironment composition calculator
CNS	Central nervous system
DNAm	DNA methylation
CpG	Cytosine-guanine dinucleotide
DMC	Differentially methylated CpG
GBM	Glioblastoma
IDH	Isocitrate dehydrogenase
TCGA	The cancer genome atlas
CPE	Consensus purity estimate
EWAS	Epigenome-wide association study
AST-HG	High-grade IDH mutant astrocytoma
AST	IDH mutant astrocytoma
OLG	Oligodendroglioma
VAMP2	Vesicle associated membrane protein 2
TREM2	Triggering receptor expressed on myeloid cells 2
HLA-DOB	Major histocompatibility complex, class II, DO beta
GABA	GABAergic neurons
GLU	Glutamatergic neurons
Oligo	Oligodendrocytes
Ast	Astrocytes
Micro	Microglia
Endo	Endothelial cells
Stromal	Stromal cells

Neu	Neutrophils
Mono	Monocytes/macrophages
NK	Natural killer cells
Treg	T regulatory cells
Bnv	Naïve B cells
Bmem	Memory B cells
CD4nv	Naïve CD4 + T cells
CD8nv	Naïve CD8 + T cells
CD4mem	Memory CD4 + T cells
CD8mem	Memory CD8 + T cells
OR	Odd's ratio
SD	Standard deviation

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s40478-024-01874-0.

Additional file 1.

Additional file 2.

Acknowledgements
S.C.P. was supported by the Burroughs Wellcome Fund Big Data in the Life Sciences Fellowship. J.K.W., A.M.M., and H.M.H. were supported by NIH grants R01 CA207360 and P50 CA097257. D.C.K. was supported by NIH grants P30 CA168524, P20 GM130423, and P20 GM103428. B.C.C. was supported by NIH grants R01CA216265, R01CA253976, P30GM149408, and P30CA023108. K.T.K. was supported by a 2018 AACR-Johnson & Johnson Lung Cancer Innovation Science (18-90-52-MICH) and NIH grant R01CA253976. Z.Z & L.A.S. was supported by CDMRP/Department of Defense (W81XWH-20-1-0778) and NIGMS (P20 GM104416). Research reported in this publication was supported by an Institutional Development Award (IDeA) from the National Institute of General Medical Sciences of the National Institutes of Health under grant number P20 GM104416. We acknowledge the following Shared Resources facilities at the Norris Cotton Cancer Center at Dartmouth with NCI Cancer Center Support Grant P30 CA023108: Immune Monitoring & Flow Cytometry.

Author contributions
S.C.P., J.K.W., Z.Z., L.A.S., A.M.M., D.C.K., B.C.C. and K.T.K. conceived the project and designed the experiments. S.C.P., Z.Z., and L.A.S. performed bioinformatic quality control and data analyses. S.C.P., J.K.W., Z.Z. and L.A.S. wrote the manuscript with input from all the co-authors. All authors read and approved the final manuscript.

Funding
National Institutes of Health,National Institute of General Medical Sciences,P20 GM104416,P30 GM149408,Brock C. Christensen,National Institutes of Health,National Cancer Institute,P30 CA023108,R01 CA207360,John K. Wiencke,P50 CA097257,P30 CA168524,P20 GM130423,R01 CA216265,Brock C. Christensen,R01 CA253976,Karl T. Kelsey,Johnson and Johnson Innovation,18-90-52-MICH.

Availability of data and materials
GIMiCC is available to download and use as an R package on GitHub (https://github.com/SalasLab/GIMiCC). The code used to perform analyses in this manuscript and available on Zenodo [106].

Declarations

Competing interests
J.K.W., K.T.K., L.A.S., and B.C.C. are co-founders of Cellintec, which had no role in this research. B.C.C. is an advisor to Guardant Health, which had no role in this research. A U.S. Provisional Patent Application entitled, "Glioma Immune Microenvironment Composition Calculator (GIMiCC) a method of estimating the proportions of eighteen key cell types from glioma DNA methylation microarray data", was filed on January 29, 2024, by Dartmouth College for the *GIMiCC* deconvolution libraries invention included in this manuscript. Inventors: Salas LA, Christensen BC, Zhang Z, Pike SC, KT Kelsey, JK Wiencke, DC Koestler, AM Molinaro. The remaining authors declare no competing interests.

Received: 16 July 2024 Accepted: 12 October 2024
Published online: 28 October 2024

References

1. Molinaro AM, Taylor JW, Wiencke JK, Wrensch MR (2019) "Genetic and molecular epidemiology of adult diffuse glioma," (in eng). Nat Rev Neurol 15(7):405–417. https://doi.org/10.1038/s41582-019-0220-2
2. Ostrom QT, Cioffi G, Waite K, Kruchko C, Barnholtz-Sloan JS (2021) "CBTRUS statistical report: primary brain and other central nervous system tumors diagnosed in the united states in 2014–2018," (in eng). Neuro Oncol 23:1–105. https://doi.org/10.1093/neuonc/noab200
3. Chen R, Smith-Cohn M, Cohen AL, Colman H (2017) "Glioma subclassifications and their clinical significance," (in eng). Neurotherapeutics 14(2):284–297. https://doi.org/10.1007/s13311-017-0519-x
4. Louis DN et al (2021) "The 2021 WHO classification of tumors of the central nervous system: a summary," (in eng). Neuro Oncol 23(8):1231–1251. https://doi.org/10.1093/neuonc/noab106
5. W. C. o. T. E. Board, *world health organization classification of tumours of the central nervous system*, 5th edition ed. international agency for research on cancer, 2021.
6. Lv B et al (2022) "Immunotherapy: reshape the tumor immune microenvironment," (in eng). Front Immunol 13:844142. https://doi.org/10.3389/fimmu.2022.844142
7. Wang XX et al (2022) "Immune gene signatures and immunotypes in immune microenvironment are associated with glioma prognose," (in eng). Front Immunol 13:823910. https://doi.org/10.3389/fimmu.2022.823910
8. Himes BT, Geiger PA, Ayasoufi K, Bhargav AG, Brown DA, Parney IF (2021) "Immunosuppression in glioblastoma: current understanding and therapeutic implications," (in eng). Front Oncol 11:770561. https://doi.org/10.3389/fonc.2021.770561
9. Burghardt I et al (2021) "Endoglin and TGF-β signaling in glioblastoma," (in eng). Cell Tissue Res 384(3):613–624. https://doi.org/10.1007/s00441-020-03323-5
10. Qian J et al (2018) "TLR2 promotes glioma immune evasion by downregulating mhc class II molecules in microglia," (in eng). Cancer Immunol Res 6(10):1220–1233. https://doi.org/10.1158/2326-6066.CIR-18-0020
11. Andersen RS, Anand A, Harwood DSL, Kristensen BW (2021) "Tumor-associated microglia and macrophages in the glioblastoma microenvironment and their implications for therapy," (in eng). Cancers 13(17):4255. https://doi.org/10.3390/cancers13174255
12. Yu MW, Quail DF (2021) "Immunotherapy for glioblastoma: current progress and challenges," (in eng). Front Immunol 12:676301. https://doi.org/10.3389/fimmu.2021.676301
13. Sener U, Ruff MW, Campian JL (2022) "Immunotherapy in glioblastoma: current approaches and future perspectives," (in eng). Int J Mol Sci 23(13):7046. https://doi.org/10.3390/ijms23137046
14. Gedeon PC et al (2020) "Checkpoint inhibitor immunotherapy for glioblastoma: current progress, challenges and future outlook," (in eng). Expert Rev Clin Pharmacol 13(10):1147–1158. https://doi.org/10.1080/17512433.2020.1817737
15. Wang Y et al (2021) "Remodelling and treatment of the blood-brain barrier in glioma," (in eng). Cancer Manag Res 13:4217–4232. https://doi.org/10.2147/CMAR.S288720
16. Toh CH, Siow TY (2021) "Factors associated with dysfunction of glymphatic system in patients with glioma," (in eng). Front Oncol 11:744318. https://doi.org/10.3389/fonc.2021.744318
17. Xu D, Zhou J, Mei H, Li H, Sun W, Xu H (2021) "Impediment of cerebrospinal fluid drainage through glymphatic system in glioma," (in eng). Front Oncol 11:790821. https://doi.org/10.3389/fonc.2021.790821
18. Gromeier M et al (2021) "Very low mutation burden is a feature of inflamed recurrent glioblastomas responsive to cancer

immunotherapy," (in eng). Nat Commun 12(1):352. https://doi.org/10.1038/s41467-020-20469-6
19. Chen Z, Hambardzumyan D (2018) "Immune microenvironment in glioblastoma subtypes," (in eng). Front Immunol 9:1004. https://doi.org/10.3389/fimmu.2018.01004
20. Abdelfattah N et al (2022) "Single-cell analysis of human glioma and immune cells identifies S100A4 as an immunotherapy target," (in eng). Nat Commun 13(1):767. https://doi.org/10.1038/s41467-022-28372-y
21. Moffet JJD et al (2023) "Spatial architecture of high-grade glioma reveals tumor heterogeneity within distinct domains," (in eng). Neurooncol Adv 5(1):142. https://doi.org/10.1093/noajnl/vdad142
22. Nicholson JG, Fine HA (2021) "Diffuse glioma heterogeneity and its therapeutic implications," (in eng). Cancer Discov 11(3):575–590. https://doi.org/10.1158/2159-8290.CD-20-1474
23. Ene CI, Holland EC (2015) "Personalized medicine for gliomas," (in eng). Surg Neurol Int 6(Suppl 1):S89-95. https://doi.org/10.4103/2152-7806.151351
24. Kiyotani K, Toyoshima Y, Nakamura Y (2021) "Personalized immunotherapy in cancer precision medicine," (in eng). Cancer Biol Med 18(4):955–965. https://doi.org/10.20892/j.issn.2095-3941.2021.0032
25. Greenberg MVC, Bourc'his D (2019) "The diverse roles of DNA methylation in mammalian development and disease," (in eng). Nat Rev Mol Cell Biol 20(10):590–607. https://doi.org/10.1038/s41580-019-0159-6
26. Farlik M et al (2016) "DNA methylation dynamics of human hematopoietic stem cell differentiation," (in eng). Cell Stem Cell 19(6):808–822. https://doi.org/10.1016/j.stem.2016.10.019
27. Okano M, Bell DW, Haber DA, Li E (1999) "DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development," (in eng). Cell 99(3):247–257. https://doi.org/10.1016/s0092-8674(00)81656-6
28. Jeong H et al (2021) "Evolution of DNA methylation in the human brain," (in eng). Nat Commun 12(1):2021. https://doi.org/10.1038/s41467-021-21917-7
29. Salas LA et al (2022) "Enhanced cell deconvolution of peripheral blood using DNA methylation for high-resolution immune profiling," (in eng). Nat Commun 13(1):761. https://doi.org/10.1038/s41467-021-27864-7
30. Zhang Z, Wiencke JK, Kelsey KT, Koestler DC, Christensen BC, Salas LA (2022) "HiTIMED: hierarchical tumor immune microenvironment epigenetic deconvolution for accurate cell type resolution in the tumor microenvironment using tumor-type-specific DNA methylation data," (in eng). J Transl Med 20(1):516. https://doi.org/10.1186/s12967-022-03736-6
31. Titus AJ, Gallimore RM, Salas LA, Christensen BC (2017) "Cell-type deconvolution from DNA methylation: a review of recent applications," (in eng). Hum Mol Genet 26(R2):R216–R224. https://doi.org/10.1093/hmg/ddx275
32. Zhang Z et al (2023) Hierarchical deconvolution for extensive cell type resolution in the human brain using DNA methylation. Front Neurosci 1698:153. https://doi.org/10.21203/rs.3.rs-2679515/v1
33. Muse ME, Carroll CD, Salas LA, Karagas MR, Christensen BC (2023) "Application of novel breast biospecimen cell-type adjustment identifies shared dna methylation alterations in breast tissue and milk with breast cancer-risk factors," (in eng). Cancer Epidemiol Biomarkers Prev 32(4):550–560. https://doi.org/10.1158/1055-9965.EPI-22-0405
34. Zhu T et al (2022) "A pan-tissue DNA methylation atlas enables in silico decomposition of human tissue methylomes at cell-type resolution," (in eng). Nat Methods 19(3):296–306. https://doi.org/10.1038/s41592-022-01412-7
35. Yang J et al (2022) "DNA methylation-based epigenetic signatures predict somatic genomic alterations in gliomas," (in eng). Nat Commun 13(1):4410. https://doi.org/10.1038/s41467-022-31827-x
36. Ferreyra Vega S, Olsson Bontell T, Corell A, Smits A, Jakola AS, Carén H (2021) "DNA methylation profiling for molecular classification of adult diffuse lower-grade gliomas," (in eng). Clin Epigenet 13(1):102. https://doi.org/10.1186/s13148-021-01085-7
37. Wu Z et al (2022) "Impact of the methylation classifier and ancillary methods on CNS tumor diagnostics," (in eng). Neuro Oncol 24(4):571–581. https://doi.org/10.1093/neuonc/noab227
38. Wenger A, Carén H (2022) "Methylation profiling in diffuse gliomas: diagnostic value and considerations," (in eng). Cancers (Basel) 14(22):5679. https://doi.org/10.3390/cancers14225679
39. Capper D et al (2018) "DNA methylation-based classification of central nervous system tumours," (in eng). Nature 555(7697):469–474. https://doi.org/10.1038/nature26000
40. Arneson D, Yang X, Wang K (2020) "MethylResolver-a method for deconvoluting bulk DNA methylation profiles into known and unknown cell contents," (in eng). Commun Biol 3(1):422. https://doi.org/10.1038/s42003-020-01146-2
41. Chakravarthy A et al (2018) "Pan-cancer deconvolution of tumour composition using DNA methylation," (in eng). Nat Commun 9(1):3220. https://doi.org/10.1038/s41467-018-05570-1
42. Singh O, Pratt D, Aldape K (2021) "Immune cell deconvolution of bulk DNA methylation data reveals an association with methylation class, key somatic alterations, and cell state in glial/glioneuronal tumors," (in eng). Acta Neuropathol Commun 9(1):148. https://doi.org/10.1186/s40478-021-01249-9
43. Weightman Potter PG et al (2021) "Attenuated induction of the unfolded protein response in adult human primary astrocytes in response to recurrent low glucose," (in eng). Front Endocrinol (Lausanne) 12:671724. https://doi.org/10.3389/fendo.2021.671724
44. Lin X et al (2018) "Cell type-specific DNA methylation in neonatal cord tissue and cord blood: a 850K-reference panel and comparison of cell types," (in eng). Epigenetics 13(9):941–958. https://doi.org/10.1080/15592294.2018.1522929
45. Kozlenkov A et al (2018) "A unique role for DNA (hydroxy)methylation in epigenetic regulation of human inhibitory neurons," (in eng). Sci Adv 4(9):eaau6190. https://doi.org/10.1126/sciadv.aau6190
46. de Witte LD et al (2022) "Contribution of age, brain region, mood disorder pathology, and interindividual factors on the methylome of human microglia," (in eng). Biol Psychiatry 91(6):572–581. https://doi.org/10.1016/j.biopsych.2021.10.020
47. Mendizabal I et al (2019) "Cell type-specific epigenetic links to schizophrenia risk in the brain," (in eng). Genome Biol 20(1):135. https://doi.org/10.1186/s13059-019-1747-7
48. Xu Z, Niu L, Li L, Taylor JA (2016) "ENmix: a novel background correction method for Illumina HumanMethylation450 BeadChip," (in eng). Nucleic Acids Res 44(3):e20. https://doi.org/10.1093/nar/gkv907
49. Xu Z, Niu L, Taylor JA (2021) "The ENmix DNA methylation analysis pipeline for Illumina BeadChip and comparisons with seven other preprocessing pipelines," (in eng). Clin Epigenet 13(1):216. https://doi.org/10.1186/s13148-021-01207-1
50. Zhou W, Triche TJ, Laird PW, Shen H (2018) "SeSAMe: reducing artifactual detection of DNA methylation by Infinium BeadChips in genomic deletions," (in eng). Nucleic Acids Res 46(20):e123. https://doi.org/10.1093/nar/gky691
51. Titus AJ, Houseman EA, Johnson KC, Christensen BC (2016) "methyLiftover: cross-platform DNA methylation data integration," (in eng). Bioinformatics 32(16):2517–2519. https://doi.org/10.1093/bioinformatics/btw180
52. Teschendorff AE et al (2013) "A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data," (in eng). Bioinformatics 29(2):189–196. https://doi.org/10.1093/bioinformatics/bts680
53. Tian Y et al (2017) "ChAMP: updated methylation analysis pipeline for Illumina BeadChips," (in eng). Bioinformatics 33(24):3982–3984. https://doi.org/10.1093/bioinformatics/btx513
54. Heiss JA, Just AC (2019) "Improved filtering of DNA methylation microarray data by detection p values and its impact on downstream analyses," (in eng). Clin Epigenet 11(1):15. https://doi.org/10.1186/s13148-019-0615-3
55. Aryee MJ et al (2014) "Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays," (in eng). Bioinformatics 30(10):1363–1369. https://doi.org/10.1093/bioinformatics/btu049
56. Triche TJ, Weisenberger DJ, Van Den Berg D, Laird PW, Siegmund KD (2013) "Low-level processing of illumina infinium DNA methylation beadarrays," (in eng). Nucleic Acids Res 41(7):e90. https://doi.org/10.1093/nar/gkt090

57. Zheng X, Zhang N, Wu HJ, Wu H (2017) "Estimating and accounting for tumor purity in the analysis of DNA methylation data from cancer studies," (in eng). Genome Biol 18(1):17. https://doi.org/10.1186/s13059-016-1143-5
58. Min JL, Hemani G, Davey Smith G, Relton C, Suderman M (2018) "Meffil: efficient normalization and analysis of very large DNA methylation datasets," (in eng). Bioinformatics 34(23):3983–3989. https://doi.org/10.1093/bioinformatics/bty476
59. Houseman EA et al (2012) "DNA methylation arrays as surrogate measures of cell mixture distribution," (in eng). BMC Bioinf 13:86. https://doi.org/10.1186/1471-2105-13-86
60. Zhou W, Laird PW, Shen H (2017) "Comprehensive characterization, annotation and innovative use of Infinium DNA methylation BeadChip probes," (in eng). Nucleic Acids Res 45(4):e22. https://doi.org/10.1093/nar/gkw967
61. Kent WJ et al (2002) "The human genome browser at UCSC," (in eng). Genome Res 12(6):996–1006. https://doi.org/10.1101/gr.229102
62. Maksimovic J, Oshlack A, Phipson B (2021) "Gene set enrichment analysis for genome-wide DNA methylation data," (in eng). Genome Biol 22(1):173. https://doi.org/10.1186/s13059-021-02388-x
63. Ritchie ME et al (2015) "limma powers differential expression analyses for RNA-sequencing and microarray studies," (in eng). Nucleic Acids Res 43(7):e47. https://doi.org/10.1093/nar/gkv007
64. Wang L, Yu CC, Liu XY, Deng XN, Tian Q, Du YJ (2021) "Epigenetic modulation of microglia function and phenotypes in neurodegenerative diseases," (in eng). Neural Plast 2021:9912686. https://doi.org/10.1155/2021/9912686
65. Shchukina I et al (2021) "Enhanced epigenetic profiling of classical human monocytes reveals a specific signature of healthy aging in the DNA methylome," (in eng). Nat Aging 1(1):124–141. https://doi.org/10.1038/s43587-020-00002-6
66. Ceccarelli M et al (2016) "Molecular profiling reveals biologically discrete subsets and pathways of progression in diffuse glioma," (in eng). Cell 164(3):550–563. https://doi.org/10.1016/j.cell.2015.12.028
67. Hussain S, Davanger S (2015) "Postsynaptic VAMP/synaptobrevin facilitates differential vesicle trafficking of GluA1 and GluA2 AMPA receptor subunits," (in eng). PLoS ONE 10(10):e0140868. https://doi.org/10.1371/journal.pone.0140868
68. Hou J, Chen Y, Grajales-Reyes G, Colonna M (2022) "TREM2 dependent and independent functions of microglia in Alzheimer's disease," (in eng). Mol Neurodegener 17(1):84. https://doi.org/10.1186/s13024-022-00588-y
69. Khantakova D, Brioschi S, Molgora M (2022) "Exploring the impact of TREM2 in tumor-associated macrophages," (in eng). Vaccines (Basel) 10:6. https://doi.org/10.3390/vaccines10060943
70. Chen X, Jensen PE (2004) "The expression of HLA-DO (H2-O) in B lymphocytes," (in eng). Immunol Res 29(1–3):19–28. https://doi.org/10.1385/IR:29:1-3:019
71. Edgar R, Tan PP, Portales-Casamar E, Pavlidis P (2014) "Meta-analysis of human methylomes reveals stably methylated sequences surrounding CpG islands associated with high gene expression," (in eng). Epigenet Chromatin 7(1):28. https://doi.org/10.1186/1756-8935-7-28
72. Aran D, Sirota M, Butte AJ (2015) "Systematic pan-cancer analysis of tumour purity," (in eng). Nat Commun 6:8971. https://doi.org/10.1038/ncomms9971
73. Yoshihara K et al (2013) "Inferring tumour purity and stromal and immune cell admixture from expression data," (in eng). Nat Commun 4:2612. https://doi.org/10.1038/ncomms3612
74. Han S et al (2020) "IDH mutation in glioma: molecular mechanisms and potential therapeutic targets," (in eng). Br J Cancer 122(11):1580–1589. https://doi.org/10.1038/s41416-020-0814-x
75. Christensen BC et al (2011) "DNA methylation, isocitrate dehydrogenase mutation, and survival in glioma," (in eng). J Natl Cancer Inst 103(2):143–153. https://doi.org/10.1093/jnci/djq497
76. Noushmehr H et al (2010) "Identification of a CpG island methylator phenotype that defines a distinct subgroup of glioma," (in eng). Cancer Cell 17(5):510–522. https://doi.org/10.1016/j.ccr.2010.03.017
77. Patel AP et al (2014) "Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma," (in eng). Science 344(6190):1396–1401. https://doi.org/10.1126/science.1254257
78. Couturier CP et al (2020) "Single-cell RNA-seq reveals that glioblastoma recapitulates a normal neurodevelopmental hierarchy," (in eng). Nat Commun 11(1):3406. https://doi.org/10.1038/s41467-020-17186-5
79. Wang Q et al (2017) "Tumor evolution of glioma-intrinsic gene expression subtypes associates with immunological changes in the microenvironment," (in eng). Cancer Cell 32(1):42-56.e6. https://doi.org/10.1016/j.ccell.2017.06.003
80. Carter SL et al (2012) "Absolute quantification of somatic DNA alterations in human cancer," (in eng). Nat Biotechnol 30(5):413–421. https://doi.org/10.1038/nbt.2203
81. Noguera-Castells A, García-Prieto CA, Álvarez-Errico D, Esteller M (2023) "Validation of the new EPIC DNA methylation microarray (900K EPIC v2) for high-throughput profiling of the human DNA methylome," (in eng). Epigenetics 18(1):2185742. https://doi.org/10.1080/15592294.2023.2185742
82. Colonna M (2023) "The biology of TREM receptors," (in eng). Nat Rev Immunol 23(9):580–594. https://doi.org/10.1038/s41577-023-00837-1
83. Nakamura K, Smyth MJ (2020) "TREM2 marks tumor-associated macrophages," (in eng). Signal Transduct Target Ther 5(1):233. https://doi.org/10.1038/s41392-020-00356-8
84. Müller S et al (2017) "Single-cell profiling of human gliomas reveals macrophage ontogeny as a basis for regional differences in macrophage activation in the tumor microenvironment," (in eng). Genome Biol 18(1):234. https://doi.org/10.1186/s13059-017-1362-4
85. Mathewson ND et al (2021) "Inhibitory CD161 receptor identified in glioma-infiltrating T cells by single-cell analysis," (in eng). Cell 184(5):1281-1298.e26. https://doi.org/10.1016/j.cell.2021.01.022
86. Johnson KC et al (2021) "Single-cell multimodal glioma analyses identify epigenetic regulators of cellular plasticity and environmental stress response," (in eng). Nat Genet 53(10):1456–1468. https://doi.org/10.1038/s41588-021-00926-8
87. Karimi E et al (2023) "Single-cell spatial immune landscapes of primary and metastatic brain tumours," (in eng). Nature 614(7948):555–563. https://doi.org/10.1038/s41586-022-05680-3
88. Hoogstrate Y et al (2023) "Transcriptome analysis reveals tumor microenvironment changes in glioblastoma," (in eng). Cancer Cell 41(4):678-692.e7. https://doi.org/10.1016/j.ccell.2023.02.019
89. Schalper KA et al (2019) "Neoadjuvant nivolumab modifies the tumor immune microenvironment in resectable glioblastoma," (in eng). Nat Med 25(3):470–476. https://doi.org/10.1038/s41591-018-0339-5
90. Pinton L et al (2019) "The immune suppressive microenvironment of human gliomas depends on the accumulation of bone marrow-derived macrophages in the center of the lesion," (in eng). J Immunother Cancer 7(1):58. https://doi.org/10.1186/s40425-019-0536-x
91. Fu W et al (2020) "Single-cell atlas reveals complexity of the immunosuppressive microenvironment of initial and recurrent glioblastoma," (in eng). Front Immunol 11:835. https://doi.org/10.3389/fimmu.2020.00835
92. Feng Q, Li L, Li M, Wang X (2020) "Immunological classification of gliomas based on immunogenomic profiling," (in eng). J Neuroinflammation 17(1):360. https://doi.org/10.1186/s12974-020-02030-w
93. Wu F et al (2023) "Immunological profiles of human oligodendrogliomas define two distinct molecular subtypes," (in eng). EBioMedicine 87:104410. https://doi.org/10.1016/j.ebiom.2022.104410
94. Yu K et al (2020) "Surveying brain tumor heterogeneity by single-cell RNA-sequencing of multi-sector biopsies," (in eng). Natl Sci Rev 7(8):1306–1318. https://doi.org/10.1093/nsr/nwaa099
95. Buonfiglioli A, Hambardzumyan D (2021) "Macrophages and microglia: the cerberus of glioblastoma," (in eng). Acta Neuropathol Commun 9(1):54. https://doi.org/10.1186/s40478-021-01156-z
96. Salas LA, Wiencke JK, Koestler DC, Zhang Z, Christensen BC, Kelsey KT (2018) "Tracing human stem cell lineage during development using DNA methylation," (in eng). Genome Res 28(9):1285–1295. https://doi.org/10.1101/gr.233213.117
97. Zhang Z, Lu Y, Vosoughi S, Levy JJ, Christensen BC, Salas LA (2023) "erarchical tumor artificial intelligence classifier traces tissue of origin and tumor type in primary and metastasized tumors using DNA

methylation," (in eng). NAR Cancer 5(2):017. https://doi.org/10.1093/narcan/zcad017
98. Andoh M, Koyama R (2021) Comparative review of microglia and monocytes in CNS phagocytosis. Cells 10:10. https://doi.org/10.3390/cells10102555
99. Laumont CM, Banville AC, Gilardi M, Hollern DP, Nelson BH (2022) "Tumour-infiltrating B cells: immunological mechanisms, clinical impact and therapeutic opportunities," (in eng). Nat Rev Cancer 22(7):414–430. https://doi.org/10.1038/s41568-022-00466-1
100. Cordell EC, Alghamri MS, Castro MG, Gutmann DH (2022) "T lymphocytes as dynamic regulators of glioma pathobiology," (in eng). Neuro Oncol 24(10):1647–1657. https://doi.org/10.1093/neuonc/noac055
101. Jaffe AE, Irizarry RA (2014) "Accounting for cellular heterogeneity is critical in epigenome-wide association studies," (in eng). Genome Biol 15(2):31. https://doi.org/10.1186/gb-2014-15-2-r31
102. Zhang Q et al (2022) "Intra-tumoral angiogenesis correlates with immune features and prognosis in glioma," (in eng). Aging (Albany NY) 14(10):4402–4424. https://doi.org/10.18632/aging.204079
103. Hu T et al (2022) "Construction and validation of an angiogenesis-related gene expression signature associated with clinical outcome and tumor immune microenvironment in glioma," (in eng). Front Genet 13:934683. https://doi.org/10.3389/fgene.2022.934683
104. Wang J, Shan A, Shi F, Zheng Q (2023) "Molecular and clinical characterization of ANG expression in gliomas and its association with tumor-related immune response," (in eng). Front Med (Lausanne) 10:1044402. https://doi.org/10.3389/fmed.2023.1044402
105. Heid IM, Küchenhoff H, Miles J, Kreienbrock L, Wichmann HE (2004) "Two dimensions of measurement error: classical and Berkson error in residential radon exposure assessment," (in eng). J Expo Anal Environ Epidemiol 14(5):365–377. https://doi.org/10.1038/sj.jea.7500332
106. Code utilized in the manuscript entitled "Glioma Immune Microenvironment Composition Calculator (GIMiCC): a method of estimating the proportions of eighteen key cell types from glioma DNA methylation microarray data" https://doi.org/10.5281/zenodo.10093224.

Publisher's Note