Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Hydroxymethylation alterations in progenitor-like cell types of pediatric central nervous system tumors are associated with cell type-specific transcriptional changes

Min Kyung Lee (min.kyung.lee.gr@dartmouth.edu)
Geisel School of Medicine at Dartmouth https://orcid.org/0000-0001-9558-4819

Nasim Azizgolshani (na2953@cumc.columbia.edu)
Columbia University Irving Medical Center

Ze Zhang (ze.zhang.gr@dartmouth.edu)
Geisel School of Medicine at Dartmouth

Laurent Perreard (Laurent.Perreard@dartmouth.edu)
Dartmouth College

Fred Kolling IV (Fred.W.Kolling.IV@dartmouth.edu)
Geisel School of Medicine at Dartmouth https://orcid.org/0000-0002-6178-9901

Lananh Nguyen (Lananh.Nguyen@unityhealth.to)
University of Toronto

George Zanazzi (george.j.zanazzi@hitchcock.org)
Geisel School of Medicine at Dartmouth

Lucas Salas (Lucas.A.Salas@dartmouth.edu)
Geisel School of Medicine at Dartmouth https://orcid.org/0000-0002-2279-4097

Brock Christensen (Brock.Christensen@dartmouth.edu)
Geisel School of Medicine at Dartmouth

Article

Keywords:

DOI: https://doi.org/

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: There is **NO** Competing Interest.

Title

Hydroxymethylation alterations in progenitor-like cell types of pediatric central nervous system tumors are associated with cell type-specific transcriptional changes

Authors & Affiliations

Min Kyung Lee [1], Nasim Azizgolshani [1,2], Ze Zhang [1], Laurent Perreard[3], Fred W. Kolling [3], Lananh N. Nguyen [4], George J. Zanazzi [3,5], Lucas A. Salas [1], Brock C. Christensen [1,6,7]

[1]Department of Epidemiology, Geisel School of Medicine at Dartmouth, Lebanon, NH, USA

[2]Department of Cardiothoracic Surgery, Columbia University Medical Center, New York, NY, USA

[3]Dartmouth Cancer Center, Geisel School of Medicine at Dartmouth, Lebanon, NH, USA

[4]Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, ON, Canada

[5]Department of Pathology and Laboratory Medicine, Geisel School of Medicine at Dartmouth, Lebanon, NH, USA

[6]Department of Molecular and Systems Biology, Geisel School of Medicine at Dartmouth, Lebanon, NH, USA

[7]Department of Community and Family Medicine, Geisel School of Medicine at Dartmouth, Lebanon, NH, USA

Corresponding author: Brock Christensen, PhD (Brock.Christensen@dartmouth.edu)

Abstract

Although intratumoral heterogeneity has been established in pediatric central nervous system tumors, epigenomic alterations at the cell type level have largely remained unresolved. To identify cell type-specific alterations to cytosine modifications in pediatric central nervous system tumors we utilized a multi-omic approach that integrated bulk DNA cytosine modification data (methylation and hydroxymethylation) with both bulk and single-cell RNA-sequencing data. We demonstrate a large reduction in the scope of significantly differentially modified cytosines in tumors when accounting for tumor cell type composition. In the progenitor-like cell types of tumors, we identified a preponderance differential CpG hydroxymethylation rather than methylation. Genes with differential hydroxymethylation, like *HDAC4* and *IGF1R,* were associated with cell type-specific changes in gene expression in tumors. Our results highlight the importance of epigenomic alterations in the progenitor-like cell types and its role in cell type-specific transcriptional regulation in pediatric CNS tumors.

Introduction

Central nervous system (CNS) tumors are the leading cause of cancer death in the pediatric population[1]. While major progress has been made in reducing the mortality in pediatric cancers in the past few decades, the magnitude of reduction in the mortality rate of CNS tumors have not been as substantial[2]. Even among patients who survive childhood cancers, those who have survived CNS tumors have the highest cumulative burden of disease post-survival[3]. Craniospinal radiation and neuro-toxic therapy are major risk factors for the future burden on quality of life with late effects including neurocognitive impairments such as academic and memory decline, and adverse health outcomes like abnormal hearing and growth hormone deficiency[4–9]. Efforts to address discrepancies in the reduction of mortality rates and extensive chronic health burdens later in life have been made with the recent advances in technology that have allowed for better insight into the molecular characterization of

pediatric CNS tumors[10–22]. Molecular biomarkers are progressively being incorporated into the diagnosis and management of certain pediatric CNS tumor types[23].

One method to supplementally diagnose and subtype CNS tumors is DNA methylation[24]. Capper et al. developed a classification method to address previous issues in inter-observer variability for histopathological diagnosis of many CNS tumors[24]. Since the development of this method, DNA methylation classification is now used regularly for certain pediatric CNS tumor types, like ependymomas, to understand the prognosis and manage treatment decisions[13,14]. This method utilizes bisulfite-treated DNA, which does not distinguish between 5-methylcytosine and 5-hydroxymethylcytosine, although it has been indicated only 5-methylcytosine signal from oxidative bisulfite-treated DNA alters the classification from this method[25,26]. Moreover, while advancements have improved management strategies for some tumor types, many other pediatric CNS tumor types remain underexplored.

DNA methylation is one of the most well-studied epigenomic marks, primarily known for its role in regulating gene expression. DNA methylation occurs when a methyl group is added to the 5-carbon position of a cytosine in the context of a Cytosine-phosphate-Guanine (CpG) dinucleotides by DNA methyltransferases (DNMTs)[27–32]. Methylation of CpG island promoters is associated with repression of gene expression while methylation of gene bodies is associated with activation of gene expression[33–35]. 5-methylcytosine (5-mC) many times co-exist with H3K9me3 marks and do not overlap with H3K4me3 marks and H2A.Z[34,36,37]. In addition, DNA methylation marks function as genome stabilizers by silencing transposable elements[34,38]. The main ways DNA methylation is altered in cancer include genome-wide hypomethylation in repetitive elements like retrotransposable elements[39,40], hypermethylation of promoters[40–43], and propensity for cytosines in CpG contexts to be mutated[44–47].

Cytosines can also remain in a hydroxymethylated state (5-hydroxymethylcytosine, 5-hmC). 5-hmC is formed when 5-mC is actively being demethylated by ten-eleven translocation (TET) enzymes[48–50]. TET enzymes add a hydroxyl group onto the methyl group to become 5-hydroxymethylcytosine, then add the hydroxyl group again to become 5-formylcytosine, then again to

become 5-carboxylcytosine, which is excised to become unmethylated[48–51]. While 5-hmC is an intermediate, it has been shown to have functional roles and be stable in the genome. Like 5-mC, 5-hmC has been associated with regulating transcription. It is enriched in gene bodies of active genes and in transcription start sites in which promoters are marked with H3K27me3 and H3K4me4[52,53]. 5-hmC has also been shown to play roles in maintaining pluripotency and tumorigenesis[52,54]. While generally 5-hmC levels are relatively much lower than 5-mC levels, higher levels of 5-hmC are found in the brain tissue compared to other tissue and in embryonal stem cells developmentally programmed neuronal cells[52,55–61]. Although progress has been made since the discovery of TET enzymes producing 5-hmC[49–51], more investigation is needed to understand the functional roles of 5-hmC. While alterations in hydroxymethylation patterns have not been as well examined, studies have indicated decreased hydroxymethylation across the genome in a variety of tumor types including adult and pediatric CNS tumors[26,54,62–70], and mutations in hydroxymethylation-associated genes such as *IDH1/2* and *TET1/2/3* have been associated with certain tumor types like gliomas and acute myeloid leukemia[62,71–74].

Numerous studies have established that brain tumors display intratumoral cellular heterogeneity[17,19,20,75–85]. While it is known that both DNA methylation and hydroxymethylation patterns are tissue type and cell type dependent[52,53,86–90], limited research has addressed cell type-specific DNA cytosine modification alterations in these tumors. This gap exists largely due to the high cost and limitations in technologies to profile cytosine modifications at the cell type-specific scale[91]. While the importance of cell type composition effects in epigenome-wide association studies has been well documented[92–96], single-cell methylation profiling strategies[97–100] are slowly developing in comparison to more accessible and commercially available genome profiling technologies focused on gene expression or chromatin accessibility. To address these shortcomings, computational methods have been developed to deconvolute cell type composition using DNA methylation for certain tissue types[91,101–109]. While these methods have greatly improved our understanding of the cell type composition effects on many epigenome-wide association studies, they have not been utilized in

investigating cell type composition effects on brain tumors due to some limited applicability in brain tissue.

In this study, we use a multi-omic approach to study cell type-level epigenomic alterations in pediatric CNS tumors to maximize the applicability of currently available methods. By integrating single nuclei RNA-seq and cytosine modification data, we provide a more complete picture of the cytosine modification alterations associated with pediatric CNS types and cytosine modifications that are associated with changes in transcription at the cell type level in pediatric CNS tumors.

Results

To assess the potential normal tissue margin in our tissues that may confound downstream analyses, we first determined the tumor purity of our pediatric CNS tumor samples that were used to measure DNA cytosine modifications. Tumor purity in our samples varied but did not significantly differ based on tumor type or grade (**Supplementary Figure 1**).

Genomic burden altered cytosine modifications

To determine the global epigenomic burden of altered cytosine modifications in pediatric CNS tumors compared to non-tumor pediatric brain tissue, we compared median beta values for both 5-hmC and 5-mC across samples at each CpG and determined the methylation dysregulation index (MDI). MDI is a summary measure of the epigenome-wide alteration of tumors compared to non-tumor tissue[110]. Tumor tissues displayed a decrease in 5-hmC and a slight increase in 5-mC compared to non-tumor tissue (**Figure 1A**). The 5-hmC MDI values were not significantly different by tumor type or by tumor grade (**Figure 1B**), whereas 5-mC MDI values varied by tumor type. Embryonal tumors had the greatest extent of epigenome-wide alteration burden compared to non-tumor tissue, astrocytomas had the lowest burden of 5-mC MDI compared to non-tumor tissue, and we observed increasing 5-mC MDI with increasing tumor grade. 5-hmC MDI and 5-mC MDI were positively correlated ($R = 0.44$, p-value = 0.013, **Figure 1C**). We repeated our analysis after removing one astrocytoma sample with an

outlier 5-hmC MDI value and observed consistent results (**Supplementary Figure 2**). We tested and confirmed that the burden of observed epigenomic alterations was not due to differences in tumor purity, (**Supplementary Figure 3, Supplementary Table 2A**). However, we did observe significant differences in 5-mC MDI by tumor grade (**Supplementary Table 2B**). While 5-hmC is prevalent at only 6% of 5-mC, the level of dysregulation of the hydroxymethylome is comparable to the level of dysregulation of the methylome with 5-hmC MDI being 49% of 5-mC MDI (**Table 2**). Our results suggest that while 5-hmC may not be as prevalent, epigenome-wide alterations of 5-hmC in tumors are occurring at comparable levels to altered 5-mC.

Cell type composition influences bulk-omics comparisons between pediatric CNS tumors and non-tumor pediatric brain tissue

We utilized our single nuclei RNA-seq data to identify the cell type composition of pediatric CNS tumor tissue and non-tumor pediatric brain tissue. Based on the cell type proportion distributions for all of our samples, we identified neuronal-like cells (NEU), neural stem cells (NSC), oligodendrocyte precursor cells (OPC), radial glial cells (RGC), and unipolar brush cells (UBC) as having the most variance (**Supplementary Figure 4A**). For each tumor type we compared proportions of cell types with non-tumor pediatric brain tissue. Supporting our principal component analysis, the cell types with the greatest differences were NEU, NSC, OPC, RGC, and UBC (**Supplementary Figure 4B**).

We conducted an epigenome-wide association study to determine the differential hydroxymethylated and methylated CpGs associated with each tumor type compared to non-tumor pediatric brain tissue. To reduce potential confounding by cell type composition, we incorporated cell type proportions as covariates in a stepwise manner to each series of linear models. Importantly, as the number of cell type proportion covariates included in the models increased, the scope of differentially hydroxymethylated and differentially methylated CpGs associated with each tumor type decreased (**Figure 2A – 2D, Supplementary Figure 5 – 8**). In addition, across our models in different tumor types, the extent of differentially hydroxymethylated CpGs (dhmCpGs) was far greater than that of

differentially methylated CpGs (dmCpGs). When all five cell types (NEU, NSC, OPC, RGC, and UBC) were incorporated into the model, we observed low number of dmCpGs associated with each tumor type. Embryonal tumors had the greatest number of dhmCpGs, and the 83.1% were specific to the embryonal tumors (**Figure 2E**). In the model with all five cell types included, 87 dhmCpGs were associated with astrocytoma, 850 dhmCpGs were associated with embryonal tumors, 31 dhmCpGs were associated with ependymoma, and 126 dhmCpGs were associated with glioneuronal/neuronal tumors. We identified 90 dhmCpGs (10.4%) that were shared across two or three of the tumor types and 28 dhmCpGs (3.2%) that were shared across all tumor types (**Figure 2E**). Our results suggest that epigenome-wide association studies comparing bulk pediatric CNS tumor tissue to non-tumor pediatric tissue are considerably influenced by the cell type composition. Moreover, it was quite unexpected that the observed differences were almost solely in hydroxymethylation and not in methylation.

We then compared transcriptome data from bulk RNA-seq in each of the tumor types with non-tumor pediatric brain tissue. The differential expression testing model included the same covariates (sex, age at diagnosis, and tumor purity) and the same five cell type proportions used for the EWAS analysis. Including proportions of major cell types of interest led to differences in an average of around 702 genes (range: 536 – 892) detected as significantly differentially expressed. In astrocytoma and glioneuronal/neuronal tumors, the adjusted model identified more genes that were significantly differentially expressed. In embryonal tumors and ependymomas, the adjusted model identified fewer genes that were significantly differentially expressed. Some key tumor progression-associated genes like *PTEN* in astrocytoma and in embryonal tumors, *MYCN* in ependymoma, and *BRCA2* in glioneuronal/neuronal tumors would not otherwise have been identified as significantly differentially expressed in the tumors had the cell type proportions not been adjusted for.

Across all tumor types, the majority of differentially expressed genes were increased in expression compared to the non-tumor pediatric brain tissue (**Supplementary Figure 9A, 10 - 13**). Almost half (43%, 3020 genes) of all genes with increased expression were shared across all tumor types (**Supplementary Figure 9B**). Among the genes with shared increases in expression in tumors

were *IRX5*, *MYOSLID*, *CWH43, ITGA2,* and *HOXA3*. Genes with increased expression across all tumor types were associated with biological oxidations and keratinization among other pathways (**Supplementary Figure 9D**). There were 253 genes (13.6%) that had decreased expression shared across tumor types (**Supplementary Figure 9C**), including *NPTXR, SCG2 , B4GAT1,* and *ATRN.* Genes that were decreased in expression across all tumor types were associated with the insulin receptor signaling and ion channel transport among other pathways (**Supplementary Figure 9E**).

To identify potentially important gene regulation by differential hydroxymethylation we compared changes in hydroxymethylation in dhmCpGs from the five-cell type-adjusted model with gene expression in each tumor type. Generally, genes with decreased hydroxymethylation levels had increased gene expression across tumor types compared to non-tumor pediatric brain tissue (**Figure 3**). Only one dhmCpGs associated with ependymoma had significant decreased expression. The dhmCpGs with differential expression did not generally favor promoters or gene body regions (**Figure 3, Supplementary Table 3**). Only embryonal tumors displayed slightly varying associations. While many of the dhmCpGs associated with embryonal tumors followed similar patterns of decreased 5-hmC levels and increased gene expression, there were some CpGs with decreased 5-hmC and decreased gene expression, as well as CpGs with increased 5-hmC with increased or decreased gene expression levels. Embryonal tumor associated dhmCpGs with significantly increased gene expression were less likely to be in promoter regions compared to dhmCpGs with significantly decreased gene expression (OR (95%CI) = 0.23 (0.064 – 0.78), p-value = 0.01). On the contrary, embryonal tumor associated dhmCpGs with significant increased expression were marginally more likely to be in gene body regions (OR (95%CI) = 2.81 (0.84 – 10.34), p-value = 0.06). We could not test for associations between promoter or gene body regions for other tumor types due to the limited number of dhmCpGs.

Interestingly, there were two CpGs with decreased 5-hmC levels and increased gene expression in astrocytoma, ependymoma, and glioneuronal/neuronal tumors: cg18280362 located in the promoter region of *CWH43* and cg08278401 located in the promoter region of *LRRC72*. In addition, we investigated the association between changes in 5-mC methylation and gene expression in the

embryonal tumors where there were 24 dmCpGs associated with significant changes in gene expression (**Supplementary Figure 14**). While we could not conduct statistical tests to test for an enrichment of promoter/gene body regions for shared dhmCpGs with increased gene expression, there were 18 dhmCpGs with increased gene expression in non-promoter regions and 3 dhmCpGs with increased gene expression in promoter regions. Moreover, there were 9 dhmCpGs with increased gene expression not in gene body regions and 12 dhmCpGs in gene body regions (**Supplementary Table 3**). Our results indicate that hydroxymethylation may be associated with changes in gene expression for certain genes in pediatric CNS tumors.

Molecular alterations in pediatric CNS tumors occur in a cell type-specific and tumor type-specific manner

One of the major questions that remains unanswered in many epigenome-wide association studies is whether altered cytosine modification can be ascribed to a specific cell type. With data from single nuclei RNA-seq for these pediatric CNS tumors and non-tumor pediatric brain tissues, we sought to identify epigenomic alterations at a cell type-specific level. To reduce the number of covariates in our analysis we focused on neuronal-like and progenitor-like cell types (**Supplementary Table 4**). The progenitor-like cells were an aggregation of neural stem cells, radial glial cells, oligodendrocyte precursor cells, and unipolar brush cells. We used an approach developed by Zheng et al[103] called CellDMC to identify cell-type-specific differentially hydroxymethylated and methylated CpGs. Using CellDMC we identified abundant dhmCpGs for each cell type and tumor type, far greater than the scope of CpGs identified with bulk tissue EWAS (**Figure 4A, Supplementary Figures 15 – 18, Supplementary Table 5**). While there were a relatively lower number of dmCpGs compared to the dhmCpGs, there were some dmCpGs detected in the cell type-specific model (**Figure 4B**). Majority of the cell type-specific dhmCpGs were tumor-type-specific (**Figure 4C – 4D, Supplementary Figure 19**). However, 128 dhmCpGs were observed in the neuronal-like cell types and 534 dhmCpGs were observed in the progenitor-like cell types across all four tumor types. While some neuronal-like cell-

specific dhmCpGs were acting on the same genes as the progenitor-like cell-specific dhmCpGs, genes that had decreased 5-hmC in the progenitor-like cells were exclusive (**Supplementary Figure 20**).

We then assessed the genomic context of cell type-specific dhmCpGs and tested for enrichment to various genomic contexts stratified by the direction of differential hydroxymethylation. Interestingly, both increased and decreased dhmCpGs in neuronal-like and progenitor-like cell types of astrocytoma and glioneuronal/neuronal tumors were enriched in similar contexts at Dnase hypersensitive sites (DHS), 1st exons, promoter regions (TSS200, TSS1500), and 5' UTR regions (**Figure 5**). dhmCpGs in ependymoma were dependent on the cell type in which it was occurring. Ependymoma associated dhmCpGs in the neuronal-like cells and CpGs with increased 5-hmC in progenitor-like cells were enriched in similar regions as the astrocytoma and glioneuronal/neuronal tumors. On the contrary, ependymoma associated CpGs with decreased 5-hmC in the progenitor-like cells were enriched in transcription factor binding sites (TFBS), 3' UTR, gene body, and exon regions. The dhmCpGs, especially for those occurring in the progenitor-like cell types, in embryonal tumors were enriched in distinct genomic contexts compared to the other tumor types. Progenitor-like cell type-specific dhmCpGs were enriched in the transcription factor binding sites, 3' UTR, gene body, exons, and enhancers.

Our findings indicate that most of the hydroxymethylation alterations occur in the progenitor-like cell types and are tumor-type-specific.

Cell type-specific gene expression changes associated with changes in hydroxymethylation

We next evaluated cell-specific gene expression changes for genes with cell-type-specific changes in hydroxymethylation. We calculated gene expression scores for genes associated with CpGs with differentially hydroxymethylated CpGs in the neuronal-like cells and progenitor-like cells for each granular cell types incorporated in our analysis for each tumor type (**Supplementary Figures 21 – 25**). Interestingly, for all tumor types, the expression scores for genes associated with CpGs with increased or decreased hydroxymethylation were increased in the oligodendrocyte precursor cells (OPCs) of the

tumors compared to non-tumor pediatric brain tissue (**Figure 6A**). Only the OPCs in embryonal tumors did not show a statistically significant increase in the expression of genes with increased 5-hmC in the progenitor-like cells. On the contrary, gene expression levels for each of the gene sets with cell type-specific alterations in 5-hmC were decreased in each of the cell types for all tumors compared to the non-tumor pediatric brain tissue.

HDAC4, established as associated with cancer progression and poor prognosis in a variety of tumor types[111–119], was one gene with cell type-specific dhmCpGs across all four tumor types. Interestingly, the majority of the CpGs with decreased 5-hmC were associated with progenitor-like cell types, while the majority of the CpGs with increased 5-hmC were associated with the neuronal-like cell types in the tumor tissue (**Figure 6B**). More than 50% of the dhmCpGs in *HDAC4* for each tumor type were in the gene body (**Table 3**). There were few dhmCpGs in the 5' UTR, TSS200, and DNase hypersensitive sites (DHS). The neuronal-like cell types had lower expression of *HDAC4* across all tumor types compared to the non-tumor tissue (**Figure 6D**). On the contrary, the progenitor-like cell types had higher levels of *HDAC4* expression.

IGF1R had dhmCpGs across all tumor types and is associated with tumorigenesis, therapy resistance, and poor survival in different cancer types, including in some pediatric CNS tumor types[120–130]. Most of the dhmCpGs with decreased 5-hmC were associated with the progenitor-like cell types in the tumor tissue while only a couple dhmCpGs were in the neuronal-like cell types of the tumor tissue (**Figure 6C**). Like *HDAC4,* the dhmCpGs in *IGF1R* were mostly located in the gene body and DNase hypersensitive sites, with a few scattered in the enhancer and 3' UTR regions (**Table 4**). Consistent with the lack of changes in hydroxymethylation in the neuronal-like cell types of the tumors, gene expression levels of *IGF1R* did not differ between tumors and the non-tumor tissue among neuronal-like cell types (**Figure 6D**). However, following the decreases in hydroxymethylation, *IGF1R* gene expression levels were higher in the progenitor-like cell types, particularly the OPCs, in the tumors than in the progenitor-like cell types of non-tumor tissue. EWAS results from bulk tumor tissue identified only

one or two CpGs in *HDAC4* and *IGF1R* as differentially hydroxymethylated in either cell type-adjusted or unadjusted model (**Table 4**).

Our results suggest potentially critical roles of hydroxymethylation of CpGs located within the gene body regions in regulating the gene expression of critical cancer genes, like *HDAC4* and *IGF1R*.

Discussion

In this study, we investigated the cell type-specific cytosine modification alterations in pediatric central nervous system tumors with a multi-omic approach. We described the cell type composition effects that occur in epigenome-wide association studies using bulk pediatric central nervous system tumors and non-tumor pediatric brain tissue. We identified that there were more differentially hydroxymethylated CpGs associated with each tumor type, particularly in the progenitor-like cell types, rather than differentially methylated CpGs. Lastly, we show that the cell type-specific changes in hydroxymethylation are associated with cell type-specific gene expression changes in pediatric central nervous system tumors.

Based on methods to classify tumor subtypes and the predominant focus on DNA methylation, it was unexpected that there were very few differentially methylated CpGs associated with each tumor type. One possible explanation for this phenomenon may be that as these are pediatric tissues, there is still ongoing development with which 5-hmC is associated. As our results suggest the epigenome-wide alterations of 5-hmC in these tumors, it may be critical to distinguish between 5-mC and 5-hmC to better understand the molecular underpinnings of these pediatric CNS tumors. Furthermore, it may be beneficial to incorporate 5-hmC into cytosine modification-based classification methods to improve performance.

Pediatric tumors are known not to have substantial genetic alterations. Our results suggest that pediatric CNS tumors may be characterized by non-mutational epigenomic reprogramming more so than genomic aberrations[131,132]. We identified a substantial number of differentially hydroxymethylated

CpGs associated with progenitor-like cell types of each tumor type. Additionally, even among the shared differentially hydroxymethylated CpGs in the progenitor-like cell types, numerous differentially hydroxymethylated CpGs were located within different genes that regulate epigenetic patterns, such as *DNMT3A, HDAC4, MLLT3,* and *KAT2B.* Furthermore, pediatric brain cancers have been shown to contain somatic mutations in epigenetic regulator genes such as *H3F3A*, *KDM6A*, and *MLL3*[133–135]. Considering the dysregulation of the epigenome may be important when developing new therapeutic strategies for these tumors.

While much more investigation has been conducted into how DNA methylation regulates gene expression, less is known about how DNA hydroxymethylation can also be associated with changes in gene expression. We identified relationships between cell type-specific hydroxymethylation patterns and cell type-specific gene expression in our pediatric CNS tumors. Our findings indicate that hydroxymethylation changes in the gene body regions can alter gene expression. Previous studies have found positive associations between DNA methylation in gene body regions and gene expression changes[33,44]. However, many genome-wide DNA methylation studies use the traditional bisulfite treatment approach to measure 5-mC. Because bisulfite treatment alone cannot distinguish between 5-mC and 5-hmC[25], some methylation signals may have been from 5-hmC. Further studies that explicitly distinguish between 5-hmC and 5-mC are needed to gain a clearer understanding of the effects of DNA cytosine modifications on gene expression.

We identified two genes, *HDAC4* and *IGF1R*, in our pediatric CNS tumors that were both epigenetically and transcriptionally altered in comparison to non-tumor pediatric brain tissue. *HDAC4* and *IGF1R* had differentially hydroxymethylated CpGs and increased expression in oligodendrocyte precursor cells across all four of our tumor types. Our results suggest a potential role of hydroxymethylation regulating genes associated with tumorigenesis. With these targets already having been studied in adult cancers, there are pharmacological inhibitors that already exist for these targets. Our study expands previously suggested ideas of targeting *HDAC4* and *IGF1R* in certain pediatric CNS tumor types[125,136,137].

Accruing a large sample size for pediatric CNS tumors is extremely difficult as they are very rare in the general population. While our study does incorporate a decent sample size for these rare tumors, the smaller sample size limited the inclusion of other variables and cell types that may affect methylation and transcription into our models. Future studies with an expanded cohort of pediatric CNS patients will allow us to assess the epigenomic alterations in additional cell types of interest, such as glial cells. Moreover, following our findings of cell type-specific changes in DNA cytosine modifications in these pediatric CNS tumors, other tumor types may also have cell type-specific that have yet to be detected. Tools to understand the cell type composition of tissues should be incorporated in bulk epigenome-wide association studies to discriminate the cell type composition effects.

Conclusion

Our study addresses gaps that currently exist in understanding epigenomic alterations at the cell type level in pediatric central nervous system tumors. Changes in hydroxymethylation were particularly drastic in progenitor-like cells and were associated with cell type level alterations in transcription. We highlight the relevance of epigenome dysregulation in pediatric central nervous system tumors that may lead us to more effective therapeutic targets.

Methods

Sample information

Cytosine modifications, bulk tissue gene expression, and single nuclei gene expression were measured in 32 pediatric CNS tumors of various types and 2 non-tumor pediatric brain tissue (**Table 1, Supplementary Table 1**). This study was approved by the Institutional Review Board Study #00030211. Only samples with all four molecular measurements were included in downstream analyses. The samples were collected from patients being treated at Dartmouth-Hitchcock Medical Center and the Dartmouth Cancer Center from 1993 to 2017. For each tumor type, the number of

samples was distributed evenly with 8 samples for astrocytoma, 6 for embryonal tumors, 10 for ependymoma, and 8 for glioneuronal/neuronal tumors. Pathological re-review for the histopathologic tumor type and grade were done according to the 2021 World Health Organization CNS tumor classification system, then categorized into broader tumor types. The non-tumor pediatric brain tissues were obtained from patients who underwent surgical resection for epilepsy.

Data collection and pre-processing

Single nuclei RNA-sequencing

The protocol to obtain single nuclei RNA-sequencing data and initial pre-processing steps were described in Chapter 3. To summarize briefly, nuclei were isolated from fresh frozen tissue samples following the Nuclei Pure Prep nuclei isolation kit (Sigma-Aldrich, St. Louis, MO). Each sample was multiplexed with lipid-tagged oligonucleotides following the MULTI-seq protocol[138]. Libraries for single nuclei RNA-seq were prepared following the 10X Genomics Single Cell Gene Expression workflows (10X Genomics, Pleasanton, CA). Libraries were pooled and sequenced using the Illumina NextSeq500 instrument. 10X Cell Ranger software was used to align sequences to the GRCh38 pre-mRNA reference genome.

Low-quality nuclei, as defined as having greater than 10,000 and less than 2,000 features and more than 5% of reads that map to mitochondrial genes, were removed for analyses. Samples were demultiplexed using an integrative approach, combining barcode based demultiplexing and genotype-based demultiplex method[139,140]. Downstream analyses for single nuclei-RNA seq were done with the Seurat package v4 in R[139,141–143].

Bulk RNA-sequencing

Unused nuclei from our single nuclei RNA-seq experiment were used for bulk RNA-sequencing. RNA was isolated following the RNeasy Plus kit (Qiagen, Hilden, Germany). Libraries for bulk RNA-seq were prepared following the Takara Pico v3 low-input protocol (Takara Bio, Kusatsu, Japan).

Quality control for raw single-end RNA-seq data was checked using FastQC v0.11.8[144]. Reads were trimmed of polyA sequences and low-quality bases using Cutadapt v2.4[145]. Reads were aligned to the human pre-mRNA genome GRCh38 with STAR v2.7.7a[146]. Quality control of aligned reads was confirmed with *CollectRNASeqMetrics* in the Picard software v2.18.29[147]. Duplicate reads were identified with *MarkDuplicates* function in the Picard software[147]. One sample with an extremely high duplicate read percentage was removed from downstream analyses. Counts per gene were estimated using the *htseq-count* function in the HTseq software v0.11.2[148].

DNA methylation and hydroxymethylation

In total, DNA from 33 paired pediatric brain tumor samples was treated with tandem bisulfite and oxidative bisulfite conversion followed by hybridization to Infinium HumanMethylationEPIC BeadChips to measure DNA methylation (5-mC) and hydroxymethylation (5-hmC). Raw BeadArray data were preprocessed using the *SeSAMe* pipeline from Bioconductor, including data normalization and quality control[149]. Cross-reactive probes, SNP-related probes, sex chromosome probes, non-CpG probes, and low-quality probes (pOOBHA > 0.05) were masked in the analysis[150]. The *oxBS.MLE* function was used to infer 5-mC and 5-hmC levels[151].

Tumor purity estimates

Tumor purity for the tissue samples with DNA cytosine modifications was estimated using the *getPurity* function with the non-tumor pediatric tumor tissue as our non-tumor reference and the low-grade glioma (LGG) option as our cancer type in the InifiniumPurify package v1.3.1 in R[152].

Statistical analyses

Epigenome-wide association studies

Linear regression models, adjusting for sex, age at diagnosis, and tumor purity in all models, were used to identify differentially methylated and hydroxymethylated CpGs associated with each tumor type compared to the non-tumor tissue. Multiple linear regression models, with adjustments for different cell type proportions identified from the single nuclei RNA-seq data, were added to the models. Linear

regression models were fit by using *lmFit* and *eBayes* functions in the limma package in R[153]. CpGs were considered differentially methylated or hydroxymethylated under the q-value threshold of 0.05.

Cell type-specific differential hydroxymethylation and methylation for each tumor type were identified using CellDMC[103]. Proportions of cell types of interest (neurons and progenitor-like cell types) were pulled from the single nuclei RNA-seq dataset. To limit overfitting the model in our relatively smaller sample size, we aggregated the progenitor-like cell types into a single cell type category. The progenitor-like cell types included neural stem cells (NSC), radial glial cells (RGC), oligodendrocyte precursor cells (OPC), and unipolar brush cells (UBC). UBCs were included due to the high levels of stemness score in the cell types identified previously.

Differential gene expression testing

Negative binomial regression models were used to identify the differential expressed genes in each tumor type compared to non-tumor tissue. One model was fit adjusting for age at diagnosis and sex. One model was fit adjusting for age at diagnosis, sex, and the proportions for cell types of interest (NEU, NSC, RGC, OPC, UBC), Negative binomial models were fit by using *DESeq* function in the DESeq2 package v1.36.0 in R[154]. Genes were considered as differentially expressed under the adjusted p-value threshold of 0.05.

Pathways enrichment testing

Reactome pathways enrichment associated with differentially expressed genes in each tumor type were identified using the *enrichPathway* function in the ReactomePA package v1.40.0 in R[155].

Genomic context enrichment test

Enrichment tests for genomic context for differentially hydroxymethylated CpGs were conducted using the Mantel-Haenszel test. The MH test was adjusted for the type of probe (Type I or Type II) used for the CpG in the Illumina Methylation EPIC array.

Availability of data and materials

The datasets used in the current study are available in GSE211362 and GSE152561.

Funding

This work was supported by a Prouty Pilot award from the Dartmouth Cancer Center and a Single-cell Pediatric Cancer Atlas (ScPCA) grant from the Alex's Lemonade Stand Foundation. MKL was supported by the Burroughs-Wellcome Fund: Big Data in the Life Sciences at Dartmouth. NA was supported by the S.M. Tenney Fellowship at Dartmouth. This work was also supported by R01CA216265, R01CA253976, and P20GM104416 – 6369 to BCC and CDMRP/Department of Defense (W81XWH-20-1-0778) and P20 GM104416-09/8299 to LAS. Single nuclei RNA-seq experiments were conducted in the Genomics and Molecular Biology Shared Resource (GMBSR) at Dartmouth, which is supported by NCI Cancer Center Support Grant 5P30CA023108 and NIH S10 (1S10OD030242) awards. Single-nuclei RNA experiments were also supported through the Dartmouth Center for Quantitative in collaboration with the GMBSR with support from NIGMS (P20GM130454) and NIH S10 (S10OD025235) awards.

Contributions

MKL, NA, and BCC designed the study. NA, GJZ, and LN identified subject populations and collected tissue samples. MKL, NA, and LP performed experiments to collect cytosine modification and gene expression data. MKL and ZZ processed data for downstream analyses. MKL performed statistical analyses under the supervision of LAS and BCC. BCC supervised the project. All authors reviewed the manuscript.

Ethics declarations

The authors declare no conflicts of interest.

References

1. Siegel, R. L., Miller, K. D., Fuchs, H. E. & Jemal, A. Cancer statistics, 2022. *Ca Cancer J Clin* 72, 7–33 (2022).

2. Smith, M. A., Altekruse, S. F., Adamson, P. C., Reaman, G. H. & Seibel, N. L. Declining childhood and adolescent cancer mortality. *Cancer* 120, 2497–2506 (2014).

3. Bhakta, N. *et al.* The cumulative burden of surviving childhood cancer: an initial report from the St Jude Lifetime Cohort Study (SJLIFE). *Lancet* 390, 2569–2582 (2017).

4. Palmer, S. L. *et al.* Patterns of Intellectual Development Among Survivors of Pediatric Medulloblastoma: A Longitudinal Analysis. *J Clin Oncol* 19, 2302–2308 (2001).

5. Robinson, K. E. *et al.* A quantitative meta-analysis of neurocognitive sequelae in survivors of pediatric brain tumors. *Pediatr Blood Cancer* 55, 525–531 (2010).

6. Ellenberg, L. *et al.* Neurocognitive Status in Long-Term Survivors of Childhood CNS Malignancies: A Report From the Childhood Cancer Survivor Study. *Neuropsychology* 23, 705–717 (2009).

7. Merchant, T. E. *et al.* Critical Combinations of Radiation Dose and Volume Predict Intelligence Quotient and Academic Achievement Scores After Craniospinal Irradiation in Children With Medulloblastoma. *Int J Radiat Oncol Biology Phys* 90, 554–561 (2014).

8. Pinto, M. D., Conklin, H. M., Li, C. & Merchant, T. E. Learning and Memory Following Conformal Radiation Therapy for Pediatric Craniopharyngioma and Low-Grade Glioma. *Int J Radiat Oncol Biology Phys* 84, e363–e369 (2012).

9. Ris, M. D. *et al.* Intellectual and academic outcome following two chemotherapy regimens and radiotherapy for average-risk medulloblastoma: COG A9961. *Pediatr Blood Cancer* 60, 1350–1357 (2013).

10. Taylor, M. D. *et al.* Molecular subgroups of medulloblastoma: the current consensus. *Acta Neuropathol* 123, 465–472 (2012).

11. Northcott, P. A. *et al.* The whole-genome landscape of medulloblastoma subtypes. *Nature* 547, 311–317 (2017).

12. Juraschka, K. & Taylor, M. D. Medulloblastoma in the age of molecular subgroups: a review: JNSPG 75th Anniversary Invited Review Article. *J Neurosurg Pediatrics* 24, 353–363 (2019).

13. Pajtler, K. W. *et al.* Molecular Classification of Ependymal Tumors across All CNS Compartments, Histopathological Grades, and Age Groups. *Cancer Cell* 27, 728–743 (2015).

14. Witt, H. *et al.* DNA methylation-based classification of ependymomas in adulthood: implications for diagnosis and treatment. *Neuro-oncology* 20, 1616–1624 (2018).

15. Jessa, S. *et al.* Stalled developmental programs at the root of pediatric brain tumors. *Nat Genet* 51, 1702–1713 (2019).

16. Zhang, L. *et al.* Single-Cell Transcriptomics in Medulloblastoma Reveals Tumor-Initiating Progenitors and Oncogenic Cascades during Tumorigenesis and Relapse. *Cancer Cell* 36, 302-318.e7 (2019).

17. Gojo, J. *et al.* Single-Cell RNA-Seq Reveals Cellular Hierarchies and Impaired Developmental Trajectories in Pediatric Ependymoma. *Cancer Cell* 38, 44-59.e9 (2020).

18. Gillen, A. E. *et al.* Single-Cell RNA Sequencing of Childhood Ependymoma Reveals Neoplastic Cell Subpopulations That Impact Molecular Classification and Etiology. *Cell Reports* 32, 108023 (2020).

19. Hovestadt, V. *et al.* Resolving medulloblastoma cellular architecture by single-cell genomics. *Nature* 572, 74–79 (2019).

20. Filbin, M. G. *et al.* Developmental and oncogenic programs in H3K27M gliomas dissected by single-cell RNA-seq. *Science* 360, 331–335 (2018).

21. Vladoiu, M. C. *et al.* Childhood cerebellar tumours mirror conserved fetal transcriptional programs. *Nature* 572, 67–73 (2019).

22. Reitman, Z. J. *et al.* Mitogenic and progenitor gene programmes in single pilocytic astrocytoma cells. *Nat Commun* 10, 3731 (2019).

23. Louis, D. N. *et al.* The 2021 WHO Classification of Tumors of the Central Nervous System: a summary. *Neuro-oncology* 23, 1231–1251 (2021).

24. Capper, D. *et al.* DNA methylation-based classification of central nervous system tumours. *Nature* 555, 469–474 (2018).

25. Booth, M. J. *et al.* Oxidative bisulfite sequencing of 5-methylcytosine and 5-hydroxymethylcytosine. *Nat Protoc* 8, 1841–1851 (2013).

26. Azizgolshani, N. *et al.* DNA 5-hydroxymethylcytosine in pediatric central nervous system tumors may impact tumor classification and is a positive prognostic marker. *Clin Epigenetics* 13, 176 (2021).

27. Sinsheimer, R. L. The action of pancreatic desoxyribonuclease: I. Isolation of mono- and dinucleotides. *J. Biol. Chem.* 208, 445–459 (1953).

28. Gold, M., Hurwitz, J. & Anders, M. The enzymatic methylation of RNA and DNA, II. on the species specificity. *Proc National Acad Sci* 50, 164–169 (1963).

29. Billen, D. & Hewitt, R. Influence of Starvation for Methionine and Other Amino Acids on Subsequent Bacterial Deoxyribonucleic Acid Replication. *J Bacteriol* 92, 609–617 (1966).

30. Billen, D. Methylation of the bacterial chromosome: an event at the “replication point”? *J Mol Biol* 31, 477–486 (1968).

31. Lark, C. Studies on the in vivo methylation of DNA in Escherichia coli 15T−. *J Mol Biol* 31, 389–399 (1968).

32. Srinivasan, P. R. & Borek, E. Enzymatic Alteration. *Science* 145, 548–553 (1964).

33. Jones, P. A. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet* 13, 484–492 (2012).

34. Petryk, N., Bultmann, S., Bartke, T. & Defossez, P.-A. Staying true to yourself: mechanisms of DNA methylation maintenance in mammals. *Nucleic Acids Res* 49, gkaa1154 (2020).

35. Ambrosi, C., Manzo, M. & Baubec, T. Dynamics and Context-Dependent Roles of DNA Methylation. *J Mol Biol* 429, 1459–1475 (2017).

36. Rose, N. R. & Klose, R. J. Understanding the relationship between DNA methylation and histone lysine methylation. *Biochimica Et Biophysica Acta Bba - Gene Regul Mech* 1839, 1362–1372 (2014).

37. Zilberman, D., Coleman-Derr, D., Ballinger, T. & Henikoff, S. Histone H2A.Z and DNA methylation are mutually antagonistic chromatin marks. *Nature* 456, 125–129 (2008).

38. Deniz, Ö., Frost, J. M. & Branco, M. R. Regulation of transposable elements by DNA modifications. *Nat Rev Genet* 20, 417–431 (2019).

39. Hansen, K. D. *et al.* Increased methylation variation in epigenetic domains across cancer types. *Nat Genet* 43, 768–775 (2011).

40. Berman, B. P. *et al.* Regions of focal DNA hypermethylation and long-range hypomethylation in colorectal cancer coincide with nuclear lamina–associated domains. *Nat Genet* 44, 40–46 (2012).

41. Moarii, M., Boeva, V., Vert, J.-P. & Reyal, F. Changes in correlation between promoter methylation and gene expression in cancer. *Bmc Genomics* 16, 873 (2015).

42. Ng, J. M.-K. & Yu, J. Promoter Hypermethylation of Tumour Suppressor Genes as Potential Biomarkers in Colorectal Cancer. *Int J Mol Sci* 16, 2472–2496 (2015).

43. Liyanage, C. *et al.* Promoter Hypermethylation of Tumor-Suppressor Genes p16INK4a, RASSF1A, TIMP3, and PCQAP/MED15 in Salivary DNA as a Quadruple Biomarker Panel for Early Detection of Oral and Oropharyngeal Cancers. *Biomol* 9, 148 (2019).

44. Baylin, S. B. & Jones, P. A. Epigenetic Determinants of Cancer. *Csh Perspect Biol* 8, a019505 (2016).

45. Pfeifer, G. P. p53 mutational spectra and the role of methylated CpG sequences. *Mutat Res Fundam Mol Mech Mutagen* 450, 155–166 (2000).

46. You, Y.-H., Li, C. & Pfeifer, G. P. Involvement of 5-methylcytosine in sunlight-induced mutagenesis. *J Mol Biol* 293, 493–503 (1999).

47. Rideout, W. M., Coetzee, G. A., Olumi, A. F. & Jones, P. A. 5-Methylcytosine as an Endogenous Mutagen in the Human LDL Receptor and p53 Genes. *Science* 249, 1288–1290 (1990).

48. Ito, S. *et al.* Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* 466, 1129–1133 (2010).

49. Ito, S. *et al.* Tet Proteins Can Convert 5-Methylcytosine to 5-Formylcytosine and 5-Carboxylcytosine. *Science* 333, 1300–1303 (2011).

50. Tahiliani, M. *et al.* Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1. *Science* 324, 930–935 (2009).

51. He, Y.-F. *et al.* Tet-Mediated Formation of 5-Carboxylcytosine and Its Excision by TDG in Mammalian DNA. *Science* 333, 1303–1307 (2011).

52. Shi, D.-Q., Ali, I., Tang, J. & Yang, W.-C. New Insights into 5hmC DNA Modification: Generation, Distribution and Function. *Frontiers Genetics* 8, 100 (2017).

53. Nestor, C. E. *et al.* Tissue type is a major modifier of the 5-hydroxymethylcytosine content of human genes. *Genome Res* 22, 467–477 (2012).

54. Thomson, J. P. & Meehan, R. R. The application of genome-wide 5-hydroxymethylcytosine studies in cancer research. *Epigenomics-uk* 9, 77–91 (2017).

55. Song, C.-X., Yi, C. & He, C. Mapping recently identified nucleotide variants in the genome and transcriptome. *Nat Biotechnol* 30, 1107–1116 (2012).

56. He, B. *et al.* Tissue-specific 5-hydroxymethylcytosine landscape of the human genome. *Nat Commun* 12, 4249 (2021).

57. Kriaucionis, S. & Heintz, N. The Nuclear DNA Base 5-Hydroxymethylcytosine Is Present in Purkinje Neurons and the Brain. *Science* 324, 929–930 (2009).

58. Kinde, B., Gabel, H. W., Gilbert, C. S., Griffith, E. C. & Greenberg, M. E. Reading the unique DNA methylation landscape of the brain: Non-CpG methylation, hydroxymethylation, and MeCP2. *Proc National Acad Sci* 112, 6800–6806 (2015).

59. Thomson, J. P. *et al.* Comparative analysis of affinity-based 5-hydroxymethylation enrichment techniques. *Nucleic Acids Res* 41, e206–e206 (2013).

60. Spada, F. *et al.* Active turnover of genomic methylcytosine in pluripotent cells. *Nat Chem Biol* 16, 1411–1419 (2020).

61. Stoyanova, E., Riad, M., Rao, A. & Heintz, N. 5-Hydroxymethylcytosine-mediated active demethylation is required for mammalian neuronal differentiation and function. *Elife* 10, e66973 (2021).

62. Jin, S.-G. *et al.* 5-Hydroxymethylcytosine Is Strongly Depleted in Human Cancers but Its Levels Do Not Correlate with IDH1 Mutations. *Cancer Res* 71, 7360–7365 (2011).

63. Liu, C. *et al.* Decrease of 5-Hydroxymethylcytosine Is Associated with Progression of Hepatocellular Carcinoma through Downregulation of TET1. *Plos One* 8, e62828 (2013).

64. Kudo, Y. *et al.* Loss of 5-hydroxymethylcytosine is accompanied with malignant cellular transformation. *Cancer Sci* 103, 670–676 (2012).

65. Lian, C. G. *et al.* Loss of 5-Hydroxymethylcytosine Is an Epigenetic Hallmark of Melanoma. *Cell* 150, 1135–1146 (2012).

66. Chen, K. *et al.* Loss of 5-hydroxymethylcytosine is linked to gene body hypermethylation in kidney cancer. *Cell Res* 26, 103–118 (2016).

67. Park, J.-L. *et al.* Decrease of 5hmC in gastric cancers is associated with TET1 silencing due to with DNA methylation and bivalent histone marks at TET1 CpG island 3′-shore. *Oncotarget* 6, 37647–37662 (2015).

68. Orr, B. A., Haffner, M. C., Nelson, W. G., Yegnasubramanian, S. & Eberhart, C. G. Decreased 5-Hydroxymethylcytosine Is Associated with Neural Progenitor Phenotype in Normal Brain and Shorter Survival in Malignant Glioma. *Plos One* 7, e41036 (2012).

69. Ficz, G. & Gribben, J. G. Loss of 5-hydroxymethylcytosine in cancer: Cause or consequence? *Genomics* 104, 352–357 (2014).

70. Johnson, K. C. *et al.* 5-Hydroxymethylcytosine localizes to enhancer elements and is associated with survival in glioblastoma patients. *Nat Commun* 7, 13177 (2016).

71. Figueroa, M. E. *et al.* Leukemic IDH1 and IDH2 Mutations Result in a Hypermethylation Phenotype, Disrupt TET2 Function, and Impair Hematopoietic Differentiation. *Cancer Cell* 18, 553–567 (2010).

72. Lu, C. *et al.* IDH mutation impairs histone demethylation and results in a block to cell differentiation. *Nature* 483, 474–478 (2012).

73. Rampal, R. *et al.* DNA Hydroxymethylation Profiling Reveals that WT1 Mutations Result in Loss of TET2 Function in Acute Myeloid Leukemia. *Cell Reports* 9, 1841–1855 (2014).

74. Duncan, C. G. *et al.* A heterozygous IDH1R132H/WT mutation induces genome-wide alterations in DNA methylation. *Genome Res* 22, 2339–2355 (2012).

75. Qazi, M. A., Bakhshinyan, D. & Singh, S. K. Deciphering brain tumor heterogeneity, one cell at a time. *Nat Med* 25, 1474–1476 (2019).

76. Sottoriva, A. *et al.* Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics. *Proc National Acad Sci* 110, 4009–4014 (2013).

77. Hoffman, M. *et al.* Intratumoral genetic and functional heterogeneity in pediatric glioblastoma. *Cancer Res* 79, canres.3441.2018 (2019).

78. Kim, E. L. *et al.* Intratumoral Heterogeneity and Longitudinal Changes in Gene Expression Predict Differential Drug Sensitivity in Newly Diagnosed and Recurrent Glioblastoma. *Cancers* 12, 520 (2020).

79. Qazi, M. A. *et al.* Intratumoral heterogeneity: pathways to treatment resistance and relapse in human glioblastoma. *Ann Oncol* 28, 1448–1456 (2017).

80. Gularyan, S. K. *et al.* Investigation of Inter- and Intratumoral Heterogeneity of Glioblastoma Using TOF-SIMS*. *Mol Cell Proteomics* 19, 960–970 (2020).

81. Patel, A. P. *et al.* Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* 344, 1396–1401 (2014).

82. Larsson, I. *et al.* Modeling glioblastoma heterogeneity as a dynamic network of cell states. *Mol Syst Biol* 17, e10105 (2021).

83. Berens, M. E. *et al.* Multiscale, multimodal analysis of tumor heterogeneity in IDH1 mutant vs wild-type diffuse gliomas. *Plos One* 14, e0219724 (2019).

84. Lopes, M. B. & Vinga, S. Tracking intratumoral heterogeneity in glioblastoma via regularized classification of single-cell RNA-Seq data. *Bmc Bioinformatics* 21, 59 (2020).

85. Lam, K. H. B., Valkanas, K., Djuric, U. & Diamandis, P. Unifying models of glioblastoma's intra-tumoral heterogeneity. *Neuro-oncology Adv* 2, vdaa096- (2020).

86. Meissner, A. *et al.* Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* 454, 766–770 (2008).

87. Sproul, D. *et al.* Tissue of origin determines cancer-associated CpG island promoter hypermethylation patterns. *Genome Biol* 13, R84 (2012).

88. Zhou, J. *et al.* Tissue-specific DNA methylation is conserved across human, mouse, and rat, and driven by primary sequence conservation. *Bmc Genomics* 18, 724 (2017).

89. Zhang, B. *et al.* Functional DNA methylation differences between tissues, cell types, and across individuals discovered using the M&M algorithm. *Genome Res* 23, 1522–1540 (2013).

90. Moss, J. *et al.* Comprehensive human cell-type methylation atlas reveals origins of circulating cell-free DNA in health and disease. *Nat Commun* 9, 5068 (2018).

91. Rahmani, E. *et al.* Cell-type-specific resolution epigenetics without the need for cell sorting or single-cell biology. *Nat Commun* 10, 3417 (2019).

92. Kim, S. *et al.* Enlarged leukocyte referent libraries can explain additional variance in blood-based epigenome-wide association studies. *Epigenomics-uk* 8, 1185–1192 (2016).

93. Jaffe, A. E. & Irizarry, R. A. Accounting for cellular heterogeneity is critical in epigenome-wide association studies. *Genome Biol* 15, R31 (2014).

94. Christensen, B. C. *et al.* Aging and Environmental Exposures Alter Tissue-Specific DNA Methylation Dependent upon CpG Island Context. *Plos Genet* 5, e1000602 (2009).

95. You, C. *et al.* A cell-type deconvolution meta-analysis of whole blood EWAS reveals lineage-specific smoking-associated DNA methylation changes. *Nat Commun* 11, 4779 (2020).

96. Reinius, L. E. *et al.* Differential DNA Methylation in Purified Human Blood Cells: Implications for Cell Lineage and Studies on Disease Susceptibility. *Plos One* 7, e41361 (2012).

97. Angermueller, C. *et al.* Parallel single-cell sequencing links transcriptional and epigenetic heterogeneity. *Nat Methods* 13, 229–232 (2016).

98. Clark, S. J., Lee, H. J., Smallwood, S. A., Kelsey, G. & Reik, W. Single-cell epigenomics: powerful new methods for understanding gene regulation and cell identity. *Genome Biol* 17, 72 (2016).

99. Schwartzman, O. & Tanay, A. Single-cell epigenomics: techniques and emerging applications. *Nat Rev Genet* 16, 716–726 (2015).

100. Smallwood, S. A. *et al.* Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. *Nat Methods* 11, 817–820 (2014).

101. Salas, L. A. *et al.* An optimized library for reference-based deconvolution of whole-blood biospecimens assayed using the Illumina HumanMethylationEPIC BeadArray. *Genome Biol* 19, 64 (2018).

102. Salas, L. A. *et al.* Enhanced cell deconvolution of peripheral blood using DNA methylation for high-resolution immune profiling. *Nat Commun* 13, 761 (2022).

103. Zheng, S. C., Breeze, C. E., Beck, S. & Teschendorff, A. E. Identification of differentially methylated cell types in epigenome-wide association studies. *Nat Methods* 15, 1059–1066 (2018).

104. Teschendorff, A. E. & Zheng, S. C. Cell-type deconvolution in epigenome-wide association studies: a review and recommendations. *Epigenomics-uk* 9, 757–768 (2017).

105. Houseman, E. A. *et al.* DNA methylation arrays as surrogate measures of cell mixture distribution. *Bmc Bioinformatics* 13, 86 (2012).

106. Rahmani, E. *et al.* BayesCCE: a Bayesian framework for estimating cell-type composition from DNA methylation without the need for methylation reference. *Genome Biol* 19, 141 (2018).

107. Rahmani, E. *et al.* Sparse PCA corrects for cell type heterogeneity in epigenome-wide association studies. *Nat Methods* 13, 443–445 (2016).

108. Waite, L. L. *et al.* Estimation of Cell-Type Composition Including T and B Cell Subtypes for Whole Blood Methylation Microarray Data. *Frontiers Genetics* 7, 23 (2016).

109. Zhang, Z. *et al.* HiTIMED: hierarchical tumor immune microenvironment epigenetic deconvolution for accurate cell type resolution in the tumor microenvironment using tumor-type-specific DNA methylation data. *J Transl Med* 20, 516 (2022).

110. O'Sullivan, D. E., Johnson, K. C., Skinner, L., Koestler, D. C. & Christensen, B. C. Epigenetic and genetic burden measures are associated with tumor characteristics in invasive breast carcinoma. *Epigenetics* 11, 344–353 (2016).

111. Gruhn, B. *et al.* The expression of histone deacetylase 4 is associated with prednisone poor-response in childhood acute lymphoblastic leukemia. *Leukemia Res* 37, 1200–1207 (2013).

112. Kang, Z.-H. *et al.* Histone Deacetylase HDAC4 Promotes Gastric Cancer SGC-7901 Cells Progression via p21 Repression. *Plos One* 9, e98894 (2014).

113. Kaowinn, S., Kaewpiboon, C., Koh, S. S., Krämer, O. H. & Chung, Y.-H. STAT1-HDAC4 signaling induces epithelial-mesenchymal transition and sphere formation of cancer cells overexpressing the oncogene, CUG2. *Oncol Rep* 40, 2619–2627 (2018).

114. Mottet, D. *et al.* HDAC4 represses p21WAF1/Cip1 expression in human cancer cells through a Sp1-dependent, p53-independent mechanism. *Oncogene* 28, 243–256 (2009).

115. Cheng, W. *et al.* HDAC4, a prognostic and chromosomal instability marker, refines the predictive value of MGMT promoter methylation. *J Neuro-oncol* 122, 303–312 (2015).

116. Cheng, C. *et al.* HDAC4 promotes nasopharyngeal carcinoma progression and serves as a therapeutic target. *Cell Death Dis* 12, 137 (2021).

117. Cai, J.-Y. *et al.* Histone deacetylase HDAC4 promotes the proliferation and invasion of glioma cells. *Int J Oncol* 53, 2758–2768 (2018).

118. Wilson, A. J. *et al.* HDAC4 Promotes Growth of Colon Cancer Cells via Repression of p21. *Mol Biol Cell* 19, 4062–4075 (2008).

119. Zeng, L.-S. *et al.* Overexpressed HDAC4 is associated with poor survival and promotes tumor progression in esophageal carcinoma. *Aging Albany Ny* 8, 1236–1248 (2016).

120. Creighton, C. J. *et al.* Insulin-Like Growth Factor-I Activates Gene Transcription Programs Strongly Associated With Poor Breast Cancer Prognosis. *J Clin Oncol* 26, 4078–4085 (2008).

121. Farabaugh, S. M., Boone, D. N. & Lee, A. V. Role of IGF1R in Breast Cancer Subtypes, Stemness, and Lineage Differentiation. *Front Endocrinol* 6, 59 (2015).

122. Maris, C. *et al.* IGF-IR: a new prognostic biomarker for human glioblastoma. *Brit J Cancer* 113, 729–737 (2015).

123. Doepfner, K. T., Spertini, O. & Arcaro, A. Autocrine insulin-like growth factor-I signaling promotes growth and survival of human acute myeloid leukemia cells via the phosphoinositide 3-kinase/Akt pathway. *Leukemia* 21, 1921–1930 (2007).

124. Chng, W. J., Gualberto, A. & Fonseca, R. IGF-1R is overexpressed in poor-prognostic subtypes of multiple myeloma. *Leukemia* 20, 174–176 (2006).

125. Svalina, M. N. *et al.* IGF1R as a Key Target in High Risk, Metastatic Medulloblastoma. *Sci Rep-uk* 6, 27012 (2016).

126. Tirrò, E. *et al.* Prognostic and Therapeutic Roles of the Insulin Growth Factor System in Glioblastoma. *Frontiers Oncol* 10, 612385 (2021).

127. Vewinger, N. *et al.* IGF1R Is a Potential New Therapeutic Target for HGNET-BCOR Brain Tumor Patients. *Int J Mol Sci* 20, 3027 (2019).

128. Zhang, Y. *et al.* Pan-Cancer Analysis of IGF-1 and IGF-1R as Potential Prognostic Biomarkers and Immunotherapy Targets. *Frontiers Oncol* 11, 755341 (2021).

129. Wang, P., Mak, V. CY. & Cheung, L. WT. Drugging IGF-1R in cancer: new insights and emerging opportunities. *Genes Dis* (2022) doi:10.1016/j.gendis.2022.03.002.

130. Hua, H., Kong, Q., Yin, J., Zhang, J. & Jiang, Y. Insulin-like growth factor receptor signaling in tumorigenesis and drug resistance: a challenge for cancer therapy. *J Hematol Oncol* 13, 64 (2020).

131. Hanahan, D. Hallmarks of Cancer: New Dimensions. *Cancer Discov* 12, 31–46 (2022).

132. Allis, C. D. & Jenuwein, T. The molecular hallmarks of epigenetic control. *Nat Rev Genet* 17, 487–500 (2016).

133. Savary, C. *et al.* Depicting the genetic architecture of pediatric cancers through an integrative gene network approach. *Sci Rep-uk* 10, 1224 (2020).

134. Huether, R. *et al.* The landscape of somatic mutations in epigenetic regulators across 1,000 paediatric cancer genomes. *Nat Commun* 5, 3630 (2014).

135. Lawlor, E. R. & Thiele, C. J. Epigenetic Changes in Pediatric Solid Tumors: Promising New Targets. *Clin Cancer Res* 18, 2768–2779 (2012).

136. Ecker, J., Witt, O. & Milde, T. Targeting of histone deacetylases in brain tumors. *Cns Oncol* 2, 359–376 (2013).

137. Bielen, A. *et al.* Enhanced Efficacy of IGF1R Inhibition in Pediatric Glioblastoma by Combinatorial Targeting of PDGFRα/β. *Mol Cancer Ther* 10, 1407–1418 (2011).

138. McGinnis, C. S. *et al.* MULTI-seq: sample multiplexing for single-cell RNA sequencing using lipid-tagged indices. *Nat Methods* 16, 619–626 (2019).

139. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* 184, 3573-3587.e29 (2021).

140. Huang, Y., McCarthy, D. J. & Stegle, O. Vireo: Bayesian demultiplexing of pooled single-cell RNA-seq data without genotype reference. *Genome Biol* 20, 273 (2019).

141. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* 177, 1888-1902.e21 (2019).

142. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* 36, 411–420 (2018).

143. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat Biotechnol* 33, 495–502 (2015).

144. Andrews, S. FastQC. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (2010).

145. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *Embnet J* 17, 10–12 (2011).

146. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21 (2013).

147. Institute, B. Picard Toolkit. https://broadinstitute.github.io/picard/ (2019).

148. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31, 166–169 (2015).

149. Zhou, W., Triche, T. J., Laird, P. W. & Shen, H. SeSAMe: reducing artifactual detection of DNA methylation by Infinium BeadChips in genomic deletions. *Nucleic Acids Res* 46, gky691- (2018).

150. Zhou, W., Laird, P. W. & Shen, H. Comprehensive characterization, annotation and innovative use of Infinium DNA methylation BeadChip probes. *Nucleic Acids Res* 45, e22–e22 (2017).

151. Xu, Z., Taylor, J. A., Leung, Y.-K., Ho, S.-M. & Niu, L. oxBS-MLE: an efficient method to estimate 5-methylcytosine and 5-hydroxymethylcytosine in paired bisulfite and oxidative bisulfite treated DNA. *Bioinformatics* 32, 3667–3669 (2016).

152. Qin, Y., Feng, H., Chen, M., Wu, H. & Zheng, X. InfiniumPurify: An R package for estimating and accounting for tumor purity in cancer methylation research. *Genes Dis* 5, 43–45 (2018).

153. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 43, e47–e47 (2015).

154. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15, 550 (2014).

155. Yu, G. & He, Q.-Y. ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization. *Mol Biosyst* 12, 477–479 (2015).

Figures



Figure 1

Figure 1 - 6

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- P2SupplementaryTables012023.pdf
- P2MainTables012023.pdf
- P2SupplementaryFigures012023.pdf
- P2SupplementaryFigures012023.pdf
- P2SupplementaryTables012023.pdf