A comparative single-cell multi-omic atlas links intrinsic regulatory programs to microenvironmental signaling in pediatric brain tumors

Yuzhe Li[1,2,3]†, Jinsong Zhang[1,2,3]†, Yifan Wei[1,2,3]†, Kaiyu Fan[4]†, Ruiyun Yang[1,2,3]†, Xu Han[4], Yanqiu Shao[5], Lei Tang[1,2,3], Weiwei Zhai[6,7], Wenze Huang[1,2,3]*, Jian Gong[4]*, Qiangfeng Cliff Zhang[1,2,3]*

[1] State Key Laboratory of Membrane Biology, Beijing Advanced Innovation Center for Structural Biology & Frontier Research Center for Biological Structure, School of Life Sciences, Tsinghua University; Beijing 100084, China.

[2]MOE Key Laboratory of Bioinformatics, Center for Life Sciences and Artificial Intelligence, Center for Synthetic and Systems Biology, School of Life Sciences, Tsinghua University; Beijing 100084, China.

[3]Tsinghua-Peking Center for Life Sciences; Beijing 100084, China.

[4]Department of Neurosurgery, Beijing Tiantan Hospital, Capital Medical University, 119 South Fourth Ring West Road, Fengtai District, Beijing, China.

[5]Xungene Therapeutics; Hangzhou 310000, Zhejiang, China.

[6]State Key Laboratory of Animal Biodiversity Conservation and Integrated Pest Management, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China.

[7]University of Chinese Academy of Sciences, Beijing 100049, China.

* Corresponding author. Email: hwz16@tsinghua.org.cn (W.H.); gongjian88@vip.163.com (J.G.); qczhang@tsinghua.edu.cn (Q.C.Z.)

†These authors contributed equally to this work.

Abstract

Pediatric brain tumors often carry few recurrent genetic alterations, placing regulatory mechanisms at the center of their pathogenesis. Yet how epigenetic states and microenvironmental cues rewire developmental programs remains poorly defined. Here we present a large-scale single-cell multi-omic atlas profiling chromatin accessibility and gene

expression simultaneously in 127,478 cells from 37 patients spanning four major pediatric brain cancer types: diffuse high-grade glioma, circumscribed low-grade glioma, medulloblastoma, and ependymoma. We identified more than 330,000 candidate cis-regulatory elements (cCREs) and 58,000 peak-gene links, reconstructing a regulatory hierarchy from cancer-type-specific enhancers to subtype-restricted domains of regulatory chromatin (DORCs). Integrating these links with transcription factor (TF) motifs and gene expression defined regulons that distinguished cancer types, molecular subtypes, and intratumoral states more effectively than TF expression or motif accessibility alone, and selected regulon activities were associated with clinical outcomes. To integrate these intrinsic programs to their microenvironmental context, we developed SIREN, a computational framework that links extracellular signals to TF regulons in recipient cells. SIREN nominated candidate immune-tumor signaling relationships associated with cancer-type- and cell-state-specific oncogenic programs. Together, this atlas provides an enabling resource for tracing pediatric brain tumor regulation from CREs to malignant cell states and candidate microenvironmental inputs, while prioritizing regulatory mechanisms for functional investigation.

Introduction

Pediatric brain tumors remain a leading cause of cancer-related mortality in children[1-3]. While current standard-of-care (*i.e.*, surgical resection followed by risk-adapted radiotherapy and chemotherapy) can achieve durable disease control in low-grade entities, it remains largely ineffective against aggressive subtypes. For instance, H3K27-altered diffuse midline gliomas carry a median survival of less than two years, and MYC-amplified Group 3 medulloblastomas (MB) have five-year overall survival of only 30-40%[4-6]. The developing brain is exquisitely vulnerable to treatment-related toxicity, leaving survivors with severe, long-term neurocognitive and developmental sequelae[7,8]. These clinical constraints underscore an urgent need for mechanistically informed, targeted therapeutic strategies.

Over the past decade, single-cell transcriptomic, genomic, and epigenomic studies[9-18] have advanced our understanding of the developmental and molecular underpinnings of these tumors. These studies have shown that malignant cells recapitulate developmental hierarchies and arrest along specific lineages[9-18], while oncohistone mutations, Polycomb dysregulation, enhancer hijacking, and other noncoding alterations converge on aberrant regulatory programs

that disrupt these trajectories [19-26]. These findings—that recurrent developmental and epigenetic dysregulation profoundly shapes pediatric brain tumors—highlight a striking and intriguing biological distinction from most adult solid tumors, in which recurrent coding alterations more frequently serve as the dominant oncogenic drivers, although epigenetic dysregulation is increasingly recognized in adult malignancies as well[27,28]. Nevertheless, existing approaches only capture complementary but disconnected regulatory layers: transcriptomic profiling defines malignant cell states without resolving the chromatin circuitry that establishes them, whereas separately measured epigenomic profiling cannot directly connect regulatory elements to transcriptional outputs. Consequently, how chromatin regulation gives rise to malignant transcriptional identity remains unresolved.

Compounding this intrinsic gap, tumor cells are embedded within a microenvironment whose extrinsic cues may influence the internal regulatory machinery to reprogram tumor-cell states[29,30]. While this tumor microenvironment is emerging as a critical prognostic determinant[31,32], how extracellular signals propagate inward to remodel specific chromatin-bound regulatory programs remains poorly understood. Existing ligand–receptor analyses generally stop at inferred communication events or downstream expression signatures[33-35], without relating extracellular signals to chromatin-defined TF regulons in recipient tumor cells. Connecting intrinsic regulatory programs with microenvironmental signaling is essential for understanding how pediatric brain tumor states are established, sustained, and remodeled.

Here, to address both gaps, we present a large-scale, comparative single-cell multi-omic atlas of chromatin accessibility and gene expression in 127,478 cells from 37 patients across four major pediatric brain cancer types: diffuse high-grade glioma (DHGG), circumscribed low-grade glioma (CLGG), MB, and ependymoma (EPN). By jointly measuring both modalities in the same cells, this resource enables systematic reconstruction of the regulatory hierarchy underlying tumor identity and cell-state heterogeneity. We mapped more than 330,000 candidate cis-regulatory elements (cCREs) and 58,000 peak–gene links, identifying subtype-restricted domains of regulatory chromatin (DORCs)[36] and super-enhancer-like hubs that distinguish molecular subtypes. We reconstructed chromatin-defined transcription factor (TF) regulons that resolved cancer types, molecular subtypes, and intratumoral states more effectively than TF expression or motif accessibility alone, with selected regulon activities associated with clinical outcomes.

Conceptually, our study was motivated by the emerging view that pediatric brain tumors are diseases of dysregulated developmental cell states rather than malignancies driven primarily by extensive genetic alterations[26,37]. Under this framework, tumor behavior is determined not only by the genetic lesions that initiate transformation but also by the regulatory circuitry that stabilizes lineage identity and by microenvironmental signals that reshape cell-state programs[17,26,38,39]. To connect these intrinsic and extrinsic layers, we introduce SIREN (SIgnal-to-REgulon Network), a computational framework that extends conventional ligand–receptor analysis by linking candidate intercellular signals to chromatin-defined TF regulons in recipient cells. Applying SIREN, we identified cancer-type- and cell-state-specific signaling–regulon associations involving EGFR, PDGFRA and other receptors, connecting candidate microenvironmental cues with oncogenic, mesenchymal, and hypoxia-associated regulons, and prioritizing a candidate VEGFA–PDGFB relationship in microglia. Together, this atlas provides an enabling resource and a regulatory blueprint for tracing pediatric brain tumor biology from CREs to malignant cell states and candidate microenvironmental inputs, while prioritizing regulatory elements, TFs, and signal-to-regulon associations for subsequent mechanistic investigation and therapeutic exploration.

Results

A comparative single-cell multi-omic atlas spanning major pediatric brain cancer types

To systematically resolve and compare the regulatory landscapes underlying pediatric brain tumors, we used Parallel-seq[40], a scalable single-cell multi-omic technology that jointly measures chromatin accessibility and gene expression in the same cell (Methods). This paired design preserves the direct relationship between regulatory chromatin and transcriptional output across heterogeneous malignant and microenvironmental populations. We profiled 37 clinical tumor samples selected to represent four major pediatric brain cancer types with distinct developmental origins and molecular drivers. This comparative design enabled us to identify core regulatory programs that are shared across cancer types and to determine how they further diversify within each type. The cohort comprised CLGG (n = 10, all pilocytic astrocytoma), DHGG (n = 9, including diffuse midline glioma, diffuse pediatric-type high-grade glioma, and diffuse hemispheric glioma), EPN (n = 5, posterior fossa and supratentorial tumor), and MB (n = 13, including sonic hedgehog (SHH), Group 3, and Group 4) (Fig. 1A-B. Supplementary Table 1).

After stringent quality control, we retained high-quality joint scATAC-and-scRNA profiles for 127,478 single cells, with a median of 1,374 unique ATAC fragments and 3,413 RNA UMIs per cell (Extended Data Fig. 1A–D), comparable to those of established joint profiling methods[36,41-45]. We integrated both modalities into a unified low-dimensional cell embedding using weighted-nearest neighbor (WNN) analysis[46] after Harmony batch correction[47] (Methods), resolving discrete clusters that mapped to major cell lineages (Fig. 1C, Extended Data Fig. 1E).

We identified three non-malignant populations: oligodendrocytes (*OPALIN*), myeloid cells (*CD68*), and lymphocytes (*CD3D*, *CD79A*). These cells clustered by cell type independently of patient of origin, confirming effective batch correction (Extended Data Fig. 1F). In contrast, putative malignant cells, marked by elevated expression of proliferation markers (*MKI67*[48], *TOP2A*[49], and *CDC6*[50]) and the stemness marker *PROM1*[51,52], separated broadly by cancer type while retaining substantial patient-level heterogeneity (Fig. 1D). This organization captured both the shared molecular identity of each cancer type and the intertumoral diversity among individual patients.

To rigorously separate malignant cells from the tumor microenvironment, we inferred copy number variations (CNVs) from both modalities, using immune cells (myeloid cells and lymphocytes) as a diploid reference (Methods). Malignant clusters exhibited widespread, patient-specific chromosomal aberrations consistent with their histopathological diagnoses, whereas immune cells showed negligible and oligodendrocytes only low-level CNV signals (Extended Data Fig. 1G-H).

Malignant clusters also retained the molecular hallmarks of their respective cancer types. For example, CLGG and DHGG clusters expressed glial-lineage and glioma genes (*e.g.* EGFR, PDGFRA)[52-54], MB clusters were characterized by the lineage TF *OTX2*[55] and the oncogene *MYCN*[56], and EPN cells by *AQP4*[57] and the lincRNA *RMST*[58] (Fig. 1E). Together, these data establish a high-resolution, comparative, multi-omic reference that captures regulatory programs shared across pediatric brain tumors alongside the cancer-type-specific programs that define their distinct malignant identities. The scale and breadth of this resource provide a foundation for identifying conserved regulatory dependencies that would remain inaccessible in studies of individual cancer types.

Pediatric brain tumors engage an enhancer-dominated cancer-associated cis-regulatory landscape

To delineate the regulatory programs governing tumor identity, we mapped cCREs by calling peaks with MACS2[59] on pseudo-bulk chromatin accessibility profiles generated separately for non-immune (malignant cells and oligodendrocytes) and immune (myeloid cells and lymphocytes) compartments (Methods). Peaks were restricted to standard chromosomes and filtered to exclude hg38 ENCODE blacklist regions[60]. In total, we identified 335,872 cCREs, including 228,423 in non-immune cells and 107,449 in immune cells. Most cCREs (80% in non-immune cells and 71% in immune cells) fell within putative enhancer regions (defined here as distal intergenic and intronic regions) (Fig. 1F), consistent with the dominant role of enhancers in cell-type-specific transcriptional regulation[61]. Notably, 16~35% of cCREs were not annotated in existing regulatory databases[62-64] (Fig. 1G). Because these catalogues are derived largely from adult and non-neural tissues and undersample the developing brain, this fraction likely reflects both incomplete reference annotation and regulatory elements specific to pediatric tumor states.

To isolate the cCREs specifically underlying tumor identity rather than tissue of origin, we identified differentially accessible cCREs (dacCREs) in each cancer type relative to its putative cell of origin[65] (Methods). We compared against developmental rather than adult-normal baselines, using cerebral astrocytes/oligodendrocytes for CLGG and DHGG, cerebral astrocytes for EPN, and cerebellar granule neuron lineages for MB[13,66,67]. This developmental referencing subtracts lineage-intrinsic accessibility, so the resulting dacCREs reflect cancer-associated regulatory changes rather than the normal programs of the cell of origin. This yielded 25,005 dacCREs with increased accessibility and 19,308 with decreased accessibility ($FDR < 0.05$). Like the full cCRE repertoire, these cancer-defining elements were enriched in regulatory regions: 60% of dacCREs (ranging from 52% to 61% across cancer types) fell in putative enhancer regions, compared to 32% in promoter regions (Fig. 1H, Extended Data Fig. 1I). The consistent enrichment of dacCREs at enhancers across all four cancer types identifies enhancer dysregulation as a shared feature of pediatric brain tumor biology, despite their distinct developmental origins.

We next asked how these dacCREs are shared across the four cancer types (Extended Data Fig. 1J). Across the four cancer types, 18,550 dacCREs (44.7%) were accessible in at least two cancer types, defining a recurrent regulatory core; the remaining 55.3% were cancer-type-

specific, consistent with the fact that dacCREs were identified relative to distinct normal cell-of-origin references for each tumor type (Extended Data Fig. 1J). These shared and cancer-type-specific elements therefore primarily capture cancer-associated regulatory alterations rather than normal lineage programmers. Together, these results establish a cancer-associated, enhancer-dominated cis-regulatory landscape in pediatric brain tumors, consisting of a shared regulatory core superimposed with extensive tumor-type-specific regulatory programs.

Pediatric brain tumors exhibit more extensive regulatory remodeling than adult glioblastoma

Because pediatric brain tumors often carry relatively few recurrent genetic alterations, regulatory rewiring may play an outsized role in their pathogenesis. We therefore asked whether the regulatory dysregulation observed in our atlas is a general feature of brain tumors or a distinctive characteristic of pediatric disease. To address this, we compared the complete set of pediatric dacCREs—pooled across all four cancer types—with adult glioblastoma (GBM) dacCREs from a published pan-cancer scATAC-seq dataset[65]. Since that study assigned putative dacCRE target genes by genomic proximity, we applied the same proximity-based strategy using Signac[60] to our dataset to enable a directly comparable analysis.

At this pan-pediatric level, brain tumors showed a greater proportion of accessible peaks classified as dacCREs than adult GBM (Fig. 1I), indicating more widespread chromatin remodeling. This difference was particularly pronounced at enhancers: 60.5% of pediatric dacCREs mapped to enhancer regions versus 55.5% in adult GBM (Fisher's exact test, $P = 2.4 \times 10^{-4}$; Extended Data Fig. 1I). Thus, the excess regulatory remodeling in pediatric brain tumors is concentrated at enhancers, potentially enabling more extensive rewiring of gene regulatory programs.

To control for histological and developmental differences, we next restricted the comparison to pediatric DHGG and adult GBM, and asked whether their distinct cis-regulatory landscapes translate into divergent target-gene programs. We identified 2,932 DHGG-associated genes and 1,096 GBM-associated genes, with 160 genes shared between the two tumor types. Pathway enrichment analysis revealed distinct regulatory programs (Extended Data Fig. 1K). DHGG-specific regulatory genes were enriched for cell-cycle progression, genome surveillance,

and proliferative transcriptional programs, including mitotic spindle, p53 pathway, DNA repair, and *MYC* targets, whereas GBM-specific regulatory genes were preferentially associated with epithelial-mesenchymal transition (EMT), hypoxia, KRAS signaling, and myogenesis. Together, these results indicate that pediatric brain tumors undergo more extensive, enhancer-centered regulatory remodeling than adult GBM, supporting regulatory rewiring as a defining feature of pediatric brain tumorigenesis.

Shared and cancer-type-specific cCRE–gene networks link dysregulated enhancers to target-gene expression

Having identified both conserved and cancer-type-specific repertoire of dysregulated enhancers, we next asked how these elements are associated with the expression of their putative target genes. We applied the FigR algorithm[68] to infer cCRE-gene associations from paired chromatin accessibility and transcriptomic profiles of malignant cells (Methods). In total we identified 58,897 candidate cCRE-gene associations involving 15,650 genes and 52,036 cCREs across the four cancer types, each gene linked to a median of four cCREs (Fig. 2A, Extended Data Fig. 2A-C). Of these, ~20% mapped to promoter regions, and a further ~20% to enhancers catalogued in GeneHancer [69] (Fig. 2B), while the remaining ~60% corresponded to candidate regulatory links not catalogued in these resources. Because GeneHancer and related catalogues are derived largely from normal adult tissues, these uncatalogued associations may include neural, developmental, and cancer-associated regulatory relationships underrepresented in existing references. The resulting network contained both recurrent cCRE-gene associations shared across cancer types and a larger set of cancer-type-specific links (Fig. 2C) . This organization is consistent with the coexistence of common and cancer-type-specific regulatory architectures reported in a recent pan-cancer multi-omics study across 11 adult cancer types[65].

Across all four pediatric brain cancer types, cCRE accessibility changes were broadly, though moderately, correlated with the expression changes of their linked genes (Pearson $\rho >$ 0.24, $P < 0.05$), indicating that chromatin accessibility and transcriptional output are coupled within these tumors (Fig. 2D). Against this backdrop of coordinated regulation, however, a subset of cCREs showed clear discordance between accessibility and expression changes, revealing a localized decoupling of epigenetic priming and transcription. The *FGFR1* and *FOXG1* loci provided two examples of such lineage-restricted, potentially poised configurations

(Fig. 2E-F, Extended Data Fig. 2D-E): whereas the oncogenic tyrosine kinase *FGFR1* was broadly expressed across all four cancer types, its linked cCRE (chr8:38401331-38402189) gained accessibility specifically in EPN (Extended Data Fig. 2F); similarly, *FOXG1*, a master regulator of forebrain development and neural progenitor identity, exhibited comparable expression across cancer types, yet its linked cCRE (chr14:28765707-28767001) specifically gained accessibility in MB (Extended Data Fig. 2G). Given the established roles of *FGFR1* and *FOXG1* in neural progenitor maintenance, lineage specification, and developmental programs[70,71], these cancer-type-specific accessibility gains are consistent with a poised regulatory state that may enable rapid or enhanced transcriptional activation in response to additional signaling cues, although this model requires functional validation.

To capture regulatory links likely rewired during tumorigenesis, we focused the cCRE-gene network analysis on dacCREs. A subset showed recurrent accessibility changes across all four cancer types, comprising both cancer-acquired gains and losses relative to non-malignant reference cells (Fig. 2E-F). For example, dacCREs linked to *BHLHE40*, a basic-helix-loop-helix protein overexpressed in gastric, breast, and brain tumors[72], gained accessibility across all four cancers, mirroring its uniform transcriptional upregulation and providing a clear example of concordant regulation. Beyond these shared programs, we also identified cancer-type-specific dacCREs nominating candidate context-specific regulatory relationships: *ETS1* in CLGG, an oncogenic TF mediating glioma cell invasion and stemness[73]; *USP22* in DHGG, which promotes glioma cell proliferation, migration, and invasion[74]; *FGFR1* in EPN, a well-established oncogenic driver of EPN[75]; and *FOXG1* in MB, a forebrain-lineage TF in MB[76]. The *FGFR1* (EPN) and *FOXG1* (MB) links correspond to the poised, cancer-type-specific elements described above, suggesting these poised chromatin states are themselves cancer-acquired regulatory features rather than lineage-inherited ones.

DORCs recover oncogenic super-enhancer programs with subtype-specific regulatory domains linked to developmental identity.

Since key developmental and oncogenic genes are often controlled by multiple regulatory elements[36], we identified DORCs, defined as higher-order regulatory domains comprising multiple cCREs linked to a common target gene. Following the framework of FigR algorithm[68], we ranked genes by their number of significant cCRE-gene associations and defined a DORC as

the full set of cCREs linked to a gene regulated by at least three such elements[36] (Fig. 3A; Methods). This yielded 984~3,186 DORCs across the four cancer types (Extended Data Fig. 3), with DORC activity defined as the aggregate accessibility of all linked cCREs.

We first asked whether DORCs recover known oncogenic super-enhancers. To address this, we intersected the constituent cCREs of each DORC with curated super-enhancer atlases[77,78] and assessed the overlap using a hypergeometric test against the background of all identified DORC cCREs (Methods). This identified DORCs that significantly overlapped cancer-type-matched super-enhancers ($FDR < 0.05$) and ranked them by enrichment (Fig. 3B-C), with the strongest hits encompassing super-enhancers at lineage- and disease-defining loci. For example, in MB, DORCs governing master regulators of the upper rhombic lip (URL) (e.g., *PAX6*)[79,80] and key determinants of photoreceptor-like lineage (*LHX4* and *PRPH2*)[81,82] densely overlapped H3K27ac-rich super-enhancers, coinciding with these core developmental programs. In total, we identified 232 MB DORCs that intersected known MB super-enhancers, and 52 EPN DORCs that overlapped known EPN super-enhancers (Fig. 3D, both $FDR < 0.05$, hypergeometric test). Curated super-enhancer atlases were not available for CLGG and DHGG, precluding an equivalent analysis in these entities.

Beyond recovering these programs, the overlapping DORCs exhibited subtype-specific activities that mirrored the developmental and oncogenic identities of each tumor subtype, suggesting that each subtype co-opts the regulatory circuitry of a distinct developmental state. In Group 4 MB, the *BARHL1* DORC was selectively active and lay in proximity to a curated MB super-enhancer (Fig. 3E); because *BARHL1* drives granule-cell specification and migration from the URL, this activity is consistent with persistence of a URL developmental program in this subgroup[83]. In Group 3 MB, the *HLX* DORC was selectively active (Fig. 3F), consistent with the role of *HLX* in Group 3 identity[84]. In SHH MB, the *MYCN* DORC was selectively active (Fig. 3G), linking the noncoding regulatory landscape to the transcriptional hyperactivation of *MYCN*, a recurrently amplified oncogene that defines this subtype[85].

A parallel pattern emerged in ependymoma. The *CACNA1H* DORC was active specifically in ST-ZFTA EPN (Fig. 3H), while the *AQP1* DORC was restricted to PFA EPN (Fig. 3I). In both cases, the DORC overlapped a curated super-enhancer at the same locus, nominating *CACNA1H* and *AQP1* as candidate regulatory genes of the ZFTA-fusion and posterior fossa group A subtypes, respectively[78,86]. Together, these analyses show that DORCs recover super-

enhancer programs and pinpoint lineage-specific developmental signatures, nominating these densely regulated loci as candidate drivers of subtype-specific transcriptional states.

Regulon landscapes resolve inter- and intra-tumoral heterogeneity and predict clinical outcomes.

Building on the cCRE-gene network, we reconstructed the upstream gene regulatory networks (GRNs) governing these programs. Specifically, we used the FigR framework to pair TFs to putative target genes by integrating (1) the enrichment of TF binding motifs within the associated cCREs and (2) TF-target co-expression across single cells (Methods). On average, each gene was regulated by 4-8 TFs and each TF by 14-35 targets (Extended Data Fig. 4A-B). We defined each "regulon" as a TF together with its inferred target genes, and scored each cell by the coordinated enrichment of target genes using the AUCell algorithm[87] (Fig. 4A, Extended Data Fig. 4C-E). Aggregated into per-patient regulon profiles, regulon scores cleanly resolved the four cancer types and their molecular subtypes (Fig. 4B). By contrast, clustering solely on average TF expression or on TF motif-derived activity scores[88] resolved these groups less clearly (ARI = 1 vs. 0.58 and 0.53; Extended Data Fig. 4F-I).

To identify candidate master regulators of each subtype, we computed regulon specificity scores (RSS)[87], ranking each regulon by how selectively it is active in one subtype relative to all other cells (Methods). Reassuringly, the top-ranked regulons recovered known subtype-defining factors: In SHH MB, these included *MEIS1*[89] and *ATOH1*[90], while Group 3 and Group 4 tumors were marked by *CRX*/*OTX2*[55,91] and *PAX6*/*BARHL1*[13,92] regulons, respectively (Fig. 4C). The same analysis resolved divergent programs demarcating the EPN PFA (*IRX2*[93], *CEBPB*[94]) and ST-ZFTA (*EMX2*[95], *TWIST1*[96]) subtypes (Extended Data Fig. 4J-K).

Regulon activity captures coordinated TF–target programs rather than individual gene expression changes. As such, it can expose intratumor heterogeneity invisible to conventional transcriptomic clustering. Indeed, clustering of regulon activity scores resolved Group 3 MB into distinct regulatory states, with two robust subtypes, G3a and G3b, clearly separated in UMAP space (Extended Data Fig. 4L), and a third cluster G3c of two patients (dominated by MB13), which was excluded from downstream analyses to avoid overinterpreting patient-specific effects. RSS analysis revealed distinct G3a- and G3b-specific regulons (Extended Data Fig. 4M), including *OTX2* and *BHLHE22*, both implicated in cancer-related processes. Projection of an

independent clinical MB cohort[97] onto our regulon framework recapitulated the major molecular groups and further resolved Group 3 tumors into G3a- and G3b-like regulatory states (Extended Data Fig. 4N). Although G3a and G3b showed only a non-significant survival trend in this limited validation cohort (Extended Data Fig. 4O), regulon activity provides a reproducible framework for resolving regulatory heterogeneity within tumor subtypes.

We next asked whether regulon activity could resolve the malignant cell states that coexist within individual tumors. DHGG offered a well-characterized test case, as its tumor cells adopt transcriptomic states analogous to those defined in adult glioblastoma[39]: astrocyte-like (AC-like), mesenchymal-like (MES-like), neural progenitor-like (NPC-like), and oligodendrocyte progenitor-like (OPC-like) , with the MES-like state, in particular, linked to aggressive progression and poor clinical outcomes[39] (Extended Data Fig. 5A). Regulon-based Leiden clustering, performed independently of state marker genes, recovered these states and revealed subpopulations governed by lineage-coherent TF programs (Fig. 4D, Extended Data Fig. 5B-C). OPC-like cells were driven by oligodendrocyte-lineage regulons (*SOX10* and *MYRF*)[98], and AC-like cells by astrocytic regulons (*PLAGL1* and *EMX2*)[99,100] (Fig. 4E, Extended Data Fig. 5D). Notably, MES-like cells exhibited selective activation of the *STAT3* regulon (RSS > 0.45), consistent with its role in mesenchymal transition and therapeutic resistance[101].

We then evaluated whether regulon activity was associated with clinical outcome by projecting the single-cell-derived regulon signatures onto bulk RNA-seq data from the independent Childhood Brain Tumor Tissue Consortium (CBTTC) cohort[102] (n = 438; AUCell scoring; Methods). In MB, high *MYC* regulon activity predicted poor overall survival (HR = 2.48, log-rank $P = 0.015$), whereas high *PAX6* activity predicted a favorable outcome (HR = 0.43, log-rank $P = 0.024$) (Fig. 4F-G). The *MYC* regulon drives effectors of cytoskeletal remodeling and cell migration, including *VCL*[103]; and *CAPNS1*[104] (Fig. 4H). Conversely, *PAX6* targets were enriched for neural-differentiation and tumor-suppressive genes, including *CHL1*[105], *RORB*[106], and *DAPK2*[107,108] (Fig. 4H). Parallel survival analyses in EPN, DHGG, and CLGG identified additional candidate prognostic regulons, reinforcing the broad translational utility of this framework (Extended Data Fig. 6A-B).

Together, these results show that chromatin-defined regulons capture the regulatory programs distinguishing cancer types, subtypes, and malignant cell states. Their association with

clinical outcome further illustrates how this atlas can translate single-cell regulatory architecture into hypotheses with potential prognostic and therapeutic relevance.

Single-cell variant analysis prioritizes candidate noncoding regulatory variants

We next asked whether somatic mutations, a major driving force in cancer, exploit the regulatory architecture defined above. Using SComatic[109], we detected candidate variants from scATAC-seq and scRNA-seq data for 25 patients with sufficient numbers of tumor and normal cells, identifying 3,505 and 73,229 mutations, respectively (the difference reflecting the distinct genomic territory and coverage of the two assays) (Fig. 5A, Extended Data Fig. 7A-B, Supplementary Table 2. Methods). Because sparse single-cell coverage and assay-specific errors complicate variant detection, we evaluated whether the candidate calls exhibited features expected of biologically relevant variants. Mutations identified from scRNA-seq data in tumor cells were skewed toward SIFT-predicted deleterious variants and enriched at evolutionarily conserved positions relative to coverage-matched control positions (Fig. 5B-C, Extended Data Fig. 7C-D), a bias support enrichment for potentially genuine, functionally consequential mutations. Consistent with this, scATAC-seq mutations were further concentrated within cCREs functionally linked to targeted genes, compared to other cCREs (19 of 25 patients with fold enrichment > 1 and $FDR < 0.05$) (Fig. 5D), an enrichment independent of sequencing read depth (Extended Data Fig. 7E). Nevertheless, these calls remain subject to accessibility, mapping quality and allelic-sampling biases and should be regarded as candidates pending orthogonal DNA-based validation.

Given their enrichment within regulatory DNA, we focused on the 696 mutations falling within cCREs confidently linked to target genes, hereafter termed cCRE-mutations (Methods). Across the four cancer types, the genes linked to cCRE-mutations were enriched for RNA splicing and chromatin remodeling pathways (Fisher's exact test, $FDR < 0.05$, Extended Data Fig. 7F). We next explored whether target-gene expression differed between mutated and non-mutated cells of the same patient (Wilcoxon rank-sum test, $P < 0.1$, $|\log2FC| > 0.5$. Fig. 5E). This analysis prioritized six of 234 tested variants, two coinciding with lower and four with higher target-gene expression. Because these associations did not remain significant after multiple-testing correction and may be influenced by cell-state composition, coverage and allele-detection biases, we interpret them as exploratory candidates rather than established regulatory effects.

For instance, a C-to-T substitution at chr20:32358104 (hg38) coincided with reduced expression of *ASXL1* (n = 163 cells, patient DMG04; log2FC = -0.89; Fig. 5F), a chromatin regulator whose loss promotes central nervous system tumorigenesis through epigenetic dysregulation[110]. In contrast, a C-to-T substitution at chr7:73522360 was observed in cells with elevated expression of *BAZ1B* (n = 326 cells, patient MB01; log2FC = 0.62; Fig. 5G), an oncogenic factor and adverse prognostic marker in gliomas[111]. These examples illustrate how the joint assay can prioritize putative regulatory variants and corresponding target genes for targeted genomic confirmation and functional testing.

Together, these findings provide an exploratory strategy for identifying candidate regulatory variants from single-cell multi-omic data. Orthogonal DNA sequencing and functional perturbation will be required to confirm their somatic origin and regulatory consequences.

SIREN links candidate microenvironmental signals to chromatin-defined regulons

The regulatory circuitry described above captures cancer-intrinsic programs but does not account for how extrinsic microenvironmental signals rewire the regulatory circuits of target cells. To address this, we developed SIREN (SIgnal-to-REgulon Network), a computational framework that integrates ligand–receptor interactions, receptor–TF relationships and context-specific regulons to nominate candidate extracellular signal-to-regulon associations (Fig. 6A, Methods). Specifically, SIREN combines ligand-receptor interactions from CellPhoneDB[112] with curated receptor-to-TF interactions from NicheNet[35], CellCall[113], and scMLnet[114], together with the regulons inferred here. generating a hypothesis-driven representation of how extracellular cues may propagate to downstream transcriptional programs.

To resolve immune-tumor signaling at fine resolution, we first clustered lymphocytes and myeloid cells by joint gene expression and chromatin accessibility profiles, defining 11 distinct cell subtypes that included seven microglial states spanning a homeostatic-to-activation continuum (distinguished by the markers in Extended Data Fig. 8). We first applied SIREN in the immune-to-tumor direction. Across four pediatric brain cancer types, we identified 2,577 (216 for CLGG, 84 for EPN, 1,043 for MB, 1,234 for DHGG; $FDR < 0.05$) significant ligand-receptor interactions linked to their downstream regulons (Supplementary Table 3). These signals were channeled prominently through a network of shared receptors, notably the receptor tyrosine

kinase (RTK) superfamily, such as ERBB4 and the atypical RTK ROR2 (Fig. 6B-C, Extended Data Fig. 9-10). The convergence of diverse microenvironmental cues onto these two key hubs is consistent with their reported involvement in driving oncogenic processes across multiple pediatric brain tumors[115,116]; notably, the downstream consequences of this shared input were subtype-specific.

In SHH MB, three ligand-receptor interactions converged on six downstream regulons (Extended Data Fig. 9A-B), most prominently *NFIA* and *NFIB*, whose target genes were enriched for hypoxia and Hedgehog signaling (FDR < 0.05), consistent with their roles in SHH MB tumorigenesis[117]. In Group 3 MB, a distinct repertoire of ligand-receptor pairs converged instead on the *OTX2*, *CRX*, and *EOMES* regulons enriched for the mesenchymal transition program (*FDR < 0.05*) (Extended Data Fig. 9C), paralleling their established oncogenic roles in this subtype[55,118].

In DHGG, receptor signaling diverged sharply across cell states, with each state engaging a distinct receptor–regulon circuit. In OPC-like cells, PDGF-A/B/C ligands engaging PDGFRA signaling activated AP-1 TFs (e.g., *FOS* and *JUN*), whose targets were enriched for the mesenchymal transition program, consistent with the association between *PDGFRA* amplification and OPC-like states[39] (Fig. 6B-C). In AC-like tumor cells, the same AP-1 regulon was activated, but predicted to be re-routed downstream of *ERBB4* rather than *PDGFRA*, which is concordant with the preferential expression of the ErbB receptor family over *PDGFRA* in AC-like cells[39] (Extended Data Fig. 10A). In MES-like tumor cells, *CD44* and *IGF1R* signaling activated the *NFE2L2* regulon, whose targets were enriched for the hypoxia pathway, consistent with the link between the mesenchymal state and hypoxic adaptation and mirroring the VEGFA-secreting MES-Hyp state in glioblastoma[119] (Extended Data Fig. 10B).

We next applied SIREN in reverse to investigate how tumor cells signal back to immune cells. Within the immune landscape, Microglia_06_IL1B, exhibited an unexpected dual phenotype: it displayed robust NF-κB-driven inflammatory activation, co-expressing *IL1B* and *NFKB1* alongside classical M1 markers[120], yet simultaneously harbored elevated pro-angiogenic signatures typically attributed to anti-inflammatory M2 programs[120] (Extended Data Fig. 8C–F). SIREN predicts that tumor-derived VEGFA signals through the NRP2 receptor on microglia to activate the *NR4A1* regulon, whose predicted downstream targets include *IL1B*, *PDGFB*, and *NLRP3*, mediators that recapitulate the Microglia_06_IL1B phenotype (Fig. 6D-E). To test this

prediction, we stimulated the human microglial cell line HMC3 (with HEK293 cells as a non-microglial control) with recombinant VEGFA and quantified *PDGFB* expression by qPCR. VEGFA stimulation upregulated *PDGFB* in HMC3, but not in HEK293 cells (fold change = 1.7, $P < 0.0001$, n = 3 biological replicates, Fig. 6F), supporting the predicted VEGFA–PDGFB axis.

Together, these findings nominate immune-derived signals that converge on shared receptors yet associate with cancer-type- and cell-state-specific tumor regulons, as well as a tumor-derived VEGFA signal linked to an inflammatory, angiogenesis-associated microglial program. More broadly, SIREN transforms the atlas from a static map of intrinsic regulation into a framework for prioritizing context-specific signal-to-regulon relationships for subsequent spatial and perturbational validation.

Discussion

Pediatric brain tumors arise from developmental programs that become arrested and subverted through epigenetic dysregulation, yet the regulatory architecture linking this intrinsic chromatin to the tumor-cell state has remained largely inaccessible. Recent single-cell studies established that these tumors recapitulate specific developmental lineages and cellular hierarchies, supporting a model in which pediatric brain tumors are fundamentally diseases of aberrantly maintained developmental cell states rather than highly mutation-driven malignancies[17,26,38,39]. By profiling chromatin accessibility and gene expression in the same127,478 cells from 37 patients spanning four major pediatric brain cancer types, we generated a comparative regulatory atlas that connects cCREs and enhancer-gene relationships with DORCs, TF regulons, malignant cell states, and candidate microenvironmental signals. Importantly, this same-cell multi-omic design enables direct reconstruction of the regulatory circuitry that may establish and stabilize developmental cell states, and that may be reshaped by both initiating genetic lesions and microenvironmental cues.

A central finding is that chromatin-defined regulon activity captures cancer type more faithfully than TF expression or motif accessibility alone. This likely reflects the fact that regulons integrate multiple layers of regulation, capturing not only TF abundance but also cofactor availability, chromatin state, and the combinatorial occupancy of CREs. Consistent with the established master-regulator paradigm in cancer[89,90], our analysis nominates *MEIS1* and *ATOH1* as candidate dependencies in SHH MB. These findings are consistent with the master-

regulator paradigm in cancer, in which lineage-specific tumor states are coordinated by a limited set of upstream regulatory factors rather than by individual downstream transcriptional targets[121-123].

The key computational advance of this work is SIREN, which connects two dimensions of tumor biology that are commonly analyzed separately: intercellular signaling and intracellular regulatory state. Existing ligand–receptor methods identify candidate communication events but generally do not specify which chromatin-defined TF programs may be engaged in recipient cells. By integrating ligand–receptor and receptor–TF priors with context-specific regulons, SIREN extends inferred communication into signal-to-regulon hypotheses. Applied across the atlas, SIREN revealed that shared receptor inputs were associated with distinct regulons depending on cancer type and malignant cell states, including PDGFRA- and ERBB4-associated programs in DHGG. In the tumor-to-immune direction, it linked VEGFA to a PDGFB-expressing microglial program, one endpoint of which was supported experimentally. SIREN therefore transforms the atlas from a map of intrinsic regulatory states into an enabling framework for prioritizing microenvironmental signals may reinforce or remodel intrinsic malignant programs, although the inferred pathways require spatial and perturbational validation.

Our paired atlas may also support exploratory prioritization of noncoding regulatory variants. Candidate cCRE variants were enriched at conserved positions and within gene-linked regulatory elements, and six showed nominal associations with target-gene expression. The *ASXL1*- and *BAZ1B*-linked variants illustrate this potential, although their somatic origin and functional consequences remain unconfirmed. These findings illustrate a feasible analytical strategy for identifying candidate noncoding alterations that may perturb cis-regulatory relationships, previously elaborated primarily in adult malignancies, such as melanoma (*TERT* promoter)[124], in pediatric brain tumors.

Several limitations should be considered. First, although our cohort of 37 patients spans four tumor entities and multiple subtypes, certain subgroups—particularly EPN subtypes and DHGG molecular variants—are represented by few samples, so subtype-specific findings should be regarded as hypothesis-generating. Second, our analyses are fundamentally correlative: regulon, noncoding-variant, and signal-to-regulon relationships are inferred from co-variation rather than established causally, and perturbation experiments—such as CRISPRi silencing of

nominated super-enhancers or regulon-defining TFs, or testing of candidate cCRE variants in patient-derived organoids—will be required to validate these dependencies.

Third, the exploratory variant analysis is constrained by the sparse coverage, allelic dropout, RNA-editing and mapping errors, and accessibility-dependent detection. The candidate variants were not corroborated by orthogonal DNA sequencing and should therefore not be interpreted as confirmed somatic or pathogenic variants. Fourth, SIREN models signaling as a linear path from ligand to receptor to TF to regulon, and does not capture the full complexity of signaling crosstalk, feedback loops, or post-translational regulation that may modulate these connections *in vivo*. Its predictions require validation through patient-level replication, spatial colocalization and perturbation of the nominated ligand, receptor, and TF components. These limitations define opportunities to extend the atlas through spatial multi-omics, larger clinical cohorts and functional perturbation studies.

In summary, this study establishes a comparative regulatory atlas spanning major pediatric brain cancer types and connects molecular layers that have previously been studied in isolation. By tracing regulatory information flow from CREs and enhancer-gene relationships to DORCs, TF regulons, malignant cell states, candidate noncoding variants, and microenvironmental signals, we propose a unified framework for understanding how developmental cell states are established, maintained, and remodeled during pediatric brain tumor progression. More broadly, our findings suggest that regulatory circuitry represents a point of convergence for both genetic and environmental influences on tumor behavior, thereby providing a conceptual framework for prioritizing candidate regulatory elements, master regulators, and signal-to-regulon pathways for future mechanistic and therapeutic investigation.

Methods

Experimental methods

Patient sample collection

Tumor samples were collected from resected specimens of 37 patients who were pathologically diagnosed with CLGG/DHGG/EPN/MB and underwent surgery at the Department of Pediatric Neurosurgery inside Beijing Tiantan Hospital, Capital Medical University. This study was approved by the Medical Research Ethics Committee of Beijing Tiantan Hospital, Capital

Medical University (No. KY2022-082-01), and written informed consent was obtained from all participants included in this study. No blinding or randomization was performed for the human tumor samples. Detailed clinical characteristics are summarized in Supplementary Table 1 and Fig. 1B, respectively.

Tumor samples preparation

Fresh tumor tissues were obtained immediately after surgical resection. Specimens were rinsed with HBSS (Gibco), kept in tissue preservation solution at 4 °C until processing. The tissue (⩽ 0.8 mg) was mechanically minced with scissors in 200 uL Buffer X (Brain Tumor Dissociation Kit, P, Miltenyi Biotec). The minced material was then transferred to a gentleMACS C Tube using a wide-bore tip. An additional 500 µL of Buffer X was used to rinse the original tube and collect residual fragments into the same C Tube, followed by addition of 50 µL Enzyme N, 3190 µL Buffer X, 40 µL Buffer Y, and 20 µL Enzyme A in the kit. The tissue was enzymatically dissociated in the C Tube using gentleMACS Octo Dissociator according to the manufacturer's instructions. The resulting suspension was filtered through a 70-µm strainer (15-1070, BIOLOGIX), washed with HBSS, and centrifuged at 300×g for 10 min at 4 °C. The pellet was resuspended in 1 mL HBSS and 3 mL RBC lysis buffer (4992957, TIANGEN) was added. The mix was incubated at room temperature for 5 min and centrifuged at 300 g for 7 min at 4 °C. Cells were resuspended in HBSS:FBS (1:1). Viable cells were counted by trypan blue exclusion. The single-cell suspension was cryopreserved in HBSS:FBS (1:1) containing 10 % DMSO, aliquoted into cryovials, frozen in a controlled-rate container at –80 °C overnight, and then transferred to liquid nitrogen.

Cryopreserved cells were rapidly thawed in a 37 °C water bath, transferred to a 15-mL tube, and centrifuged at 500×g for 5 min at 4 °C. The supernatant was discarded, and the pellet was resuspended in 3.1 mL ice-cold PBS; 0.9 mL of Debris Removal Solution (Miltenyi Biotec) was added and mixed thoroughly. To remove debris, 4 mL PBS was added dropwise to create an interface, and the tube was centrifuged at 3000×g for 10 min at 4 °C. After removing 7 mL of the upper phase, the remaining suspension was diluted with 12 mL cold PBS, inverted to resuspend, and centrifuged at 1000×g for 10 min at 4 °C. The supernatant was aspirated completely, and the pellet was resuspended in 110 µL Annexin V Binding Buffer and transferred to a 1.7-mL tube. Cells were blocked with 5 µL Human TruStain FcX (422302, BioLegend) for 5 min at room

temperature, then stained with 5 µL APC Annexin V (640941, BioLegend) and 5 µL anti-CD45-PE (304039, BioLegend) antibodies, incubated on ice in the dark for 15 min (with gentle mixing midway), and washed with 1 mL HBSS containing 5 % BSA by centrifugation at 500×g for 5 min at 4 °C. The pellet was resuspended in 1 mL of 10 µM Calcein-AM (425201, BioLegend) (prepared in HBSS with 5 % BSA from a 1 M DMSO stock) and incubated at room temperature in the dark for 20 min. After centrifugation at 500×g for 5 min at 4 °C, cells were resuspended in 500 µL HBSS with 2 % FBS, stained with 5 µL 7-AAD Viability Staining Solution (420404, BioLegend), and kept on ice in the dark for at least 5 min. Finally, the suspension was filtered through a cell strainer (pre-wetted with HBSS) fitted on a 5-mL round-bottom polypropylene tube, and the strainer was rinsed with 200 µL HBSS containing 2 % FBS to collect residual cells prior to single-cell sorting. Single cells stained as Calcein-AM positive, 7-AAD negative, and Annexin V negative were sorted with MoFlo Astrios EQ Cell Sorter (Beckman Coulter).

Parallel-split-seq analysis

Parallel-split-seq were performed as previously described with some adaptions[40]. Detailed procedure were described below.

Transposon preparation

Tn5MErev, Tn5ME-A, and barcoded Tn5ME-R1BC were ordered from Sangon Biotech (Shanghai) Co., Ltd. And sequences are listed in Supplementary Table 4. Tn5MErev (10 mM) was annealed with 10 mM Tn5ME-A and 10 mM Tn5ME-R1BC of each barcode separately with the protocol: denaturing at 95 C for 2 min, cool to 20 C with ramp rate of 0.1 C/s, and hold at 4 C. To assemble the barcoded transposons, we combined 2 ml annealed Tn5ME-A, 2 ml annealed Tn5ME-R1BC, 2 ml 10x TPS, 4 ml Tranposase (M0221, Robustnique), and 10 ml UltraPure DNase/RNase-Free Distilled Water (10977023, ThermoFisher) and incubated the mixture at room temperature for 30 min. Assembled transposons were split into 4 ml in each tube for storage no more than one month at -20 °C.

Fixation

We counted 50k sorted single cells as the starting materials for each tube. Single cells were centrifuged at 500 g for 5 min at 4 °C and resuspended with 100 µl PBS. Fixations were performed by adding 300 µl PBS containing 1.33% methanol-free formaldehyde (28906, ThermoFisher) and incubating on ice for 10 min. The fixed cells were collected by swinging-bucket centrifugation at 1000 g for 3 min at 4 °C followed by fixed-angle centrifugation at 1000 g for 1 min.

Permeabilization

A 50-ml stock of 2x RSB buffer was generated as described in Omni-ATAC[125] by combining 1 ml of 1M Tris-HCl pH 7.4 (T2663-1L, Sigma Aldrich), 200 µl of 5M NaCl (AM9759, ThermoFisher), 300 µl of 1M $MgCl_2$ (AM9530G, ThermoFisher), and 48.5 ml of UltraPure Dnase/RNase-Free Distilled Water. For each sample, a permeabilization buffer was freshly assembled from 50 µl of 2x RSB buffer, 1 µl of RiboLock, 1 µl of SUPERase•In RNase Inhibitor, 1 µl of 10% Nonidet P40 Substitute (11332473001, Sigma Aldrich), 1 µl of 10% Tween 20 (11332465001, Sigma Aldrich), 1 µl of 1% Digitonin (D141-100MG, Sigma Aldrich), 5 µl of 20% BSA, and 40 µl of UltraPure Dnase/RNase-Free Distilled Water. A Wash buffer was also prepared per sample by combining 500 µl of 2 x RSB buffer, 10 µl of 10% Tween 20, 1 µl of RiboLock, 50 µl of 20% BSA, and 439 µl of UltraPure Dnase/RNase-Free Distilled Water. Immediately after discarding the fixation solution, 100 µl permeabilization buffer was added to each tube, and the contents were pipetted up and down eight times, then placed on ice for 5 min. After permeabilization, 1 ml wash buffer was added into tubes of each sample. The cells were collected by centrifugation.

Transposition

For each sample, we prepared the ATAC-seq reaction solution by combining 10 µl of 5x LM buffer (M0221, Robustnique), 16.5 µl of PBS, 0.5 µl of RiboLock, 0.5 µl of SUPERase•In RNase Inhibitor, 0.5 µl of 10% Tween 20, 0.5 µl of 1% Digitonin, and 17.5 µl of UltraPure Dnase/RNase-Free Distilled Water, yielding a total of 46 µL. The permeabilized single cells were resuspended with the 46-µl ATAC-seq reaction solution, and then 4 µl barcoded transposon (BC1) was added to each tube, bringing the final reaction volume to 50 µL. The transposition reaction was carried out at 37 °C at 800 r.p.m. with a heated lid on the

Thermomixer. The ATAC-seq reaction were terminated by adding 2 µl of 500 µM EDTA (AM9260G, ThermoFisher). Cells were rinsed with a washing solution composed of 949 µl of PBS, 10 µl of 10% Triton X-100 (93443, Sigma Aldrich), 1 µl of RiboLock, and 50 µl of 20% BSA and collected by centrifugation.

In-cell reverse transcription

For each sample, we assembled a 16-µl resuspension cocktail containing8 µl of PBS, 0.5 µl of RiboLock, 0.5 µl of SUPERase•In RNase Inhibitor, and 7 µl of UltraPure Dnase/RNase-Free Distilled Water. The reverse transcription mixture was formulated from 8 µl of 5 x RT buffer, 2 µl of 10 mM dNTP (N0447L, NEB), 0.5 µl of RiboLock, 0.25 µl of SUPERase•In RNase Inhibitor, 4 µl of Maxima H Minus Reverse Transcriptase (EP0753, ThermoFisher), and 5.25 µl of UltraPure Dnase/RNase-Free Distilled Water. A 20-µl aliquot of this reverse transcription mixture was distributed into each tube and the 2 µl random primers and 2 µl polyT primers (each carrying the matching BC1 barcode were added into corresponding tubes (Supplementary Table 4). The transposed cells were resuspended with 16 µl resuspension cocktail and transferred to the tubes pre-loaded with the barcode-specific primers. The reaction was mixed thoroughly and the reverse transcription reaction was conducted at 50 °C for 10 minutes, followed by three thermal cycles (8 °C for 12 s, 15 °C for 45 s, 20 °C for 45 s, 30 °C for 30 s, 42 °C for 120 s, and 50 °C for 180 s), then incubating at 50 °C for 5 minutes and finally holding at 4 °C. The reverse transcription reactions from different tubes were combined into one 1.5 ml tube on ice. Cells were centrifuged and washed with 1 ml PBS supplemented with 10 µl 10% Triton X-100.

Preparation of ligation plates

We used a ligation reaction to add second and third round barcodes (Barcode2 and Barcode3). Prior to the in-cell barcode ligation, the adapters were pre-annealed in a 96-well plate. For Barcode2, 11 µM of Round2-linker was mixed with 12 µM of the corresponding barcoded Round2-BC oligonucleotides; for Barcode3, 13 µM of Round3-linker was combined with 14 µM of barcoded Round3-BC oligos (Supplementary Table 4), resulting in a final reaction volume of 100 ml per well. The plate was then incubated at 95 °C for 2 minutes, and subsequently cooled down to 20 °C at the ramp rate of 0.1 °C/s. After annealing, the entire plate content was divided

among ten separate ligation plates, with each well receiving 10 µL of the resulting barcoded splint ligation adapter mixture.

In-cell ligation

To incorporate the two barcode rounds, the ligation was performed twice following the SPLiT-seq protocol[126] with RNase inhibitor excluded from both reactions. In each round, we set up 2 ml of 1x NEBuffer 3.1 (B7203S, NEB) and 2 ml of ligation master mix (500 µl of 10x T4 DNA ligation buffer, 100 µl of T4 DNA ligase (M0082, Robustnique), 50 µl of 10% Triton X-100, and 1,350 µl of UltraPure Dnase/RNase-Free Distilled Water). The pooled single cells were re-suspended in the 1x NEBbuffer 3.1 and thoroughly mixed with the ligation master mix. Aliquots of 40 µl of this cell-ligation mixture were transferred into each well of the pre-prepared ligation plate. The plate was incubated at ambient temperature with gentle rotation at 15 rpm for 1 hour. After incubation, 2 µl of 500 µM EDTA was dispensed into each well to stop the reaction, and the contents from all wells were combined into a single tube. The pooled cells were supplemented with 50 µl of 10% Triton X-100 and centrifuged to remove supernatant. Then the cells were subjected to an additional rinse with 940 µl of PBS and 10 µl of 10% Triton X-100.

RNase digestion

Following barcode ligation, the cells were processed for RNase digestion. A digestion mixture was prepared by combining 40 µl of 5 x RT buffer, 8 µl of RNase Cocktail Enzyme Mix (AM2286, ThermoFisher), 8 µl of RNAse H (Y9220L, Enzymatics), and 144 µl of UltraPure Dnase/RNase-Free Distilled Water. The cell pellet was resuspended in this reaction mix and incubated at 37 °C for 30 min with intermittent agitation (15s shaking at 300 rpm followed by 45s rest) on the thermomixer. After digestion, cells were rinsed by adding 790 µl of PBS and 10 µl of 10% Triton X-100 then pelleted by centrifugation.

Second strand synthesis

Following RNase digestion, the pelleted cells were taken up in a second strand synthesis master mix containing 40 µl of 5 x RT buffer, 48 µl of 50% PEG 8000 (B1004SVIAL, NEB), 20 µl of 10mM dNTP, 2 µl of 1mM dN-P5 short primer (Supplementary Table 4), 5 µl of Klenow fragment (M0212L, NEB), and 85 µl of UltraPure Dnase/RNase-Free Distilled Water. The

suspension was incubated at 37°C for 1 hour with intermittent agitation (15s shaking at 300 rpm followed by 45s rest) on the thermomixer. After second strand synthesis, the cells were rinsed twice with PBS containing 0.1% Triton X-100, then re-suspended in 40 µl 0.5x PBS. An aliquot was stained with trypan blue and counted using a C-Chip disposable hemocytometer. The concentration was adjusted with additional 0.5× PBS to ⩽1500 cells/µL, and 2.5-µL portions were distributed into 0.2-mL PCR tubes. To each tube, 2.5 µL of lysis buffer (100 mM Tris-HCl pH 8.0, 1% Tween-20, 1% IGPAL CA-630, 4 µg/µL Qiagen Protease) was added and mixed. The tubes were then placed in a thermal cycler and kept at 55 °C for 8 h, immediately heated to 70 °C for 6 min to inactivate protease, and finally cooled to 4 °C. For long-term storage, the lysates were kept at –80 °C.

libraries construction and sequencing

Pre-amplification was performed by adding 12.5 µL NEBNext High-Fidelity 2× PCR Master Mix, 2.5 µL P7-index, 1.25 µL N5-index, 1.25 µL P5-solexa primer, and 3 µL water to each cell lysate (total 25 µL). The PCR program was: 72 °C for 5 min; 98 °C for 3 min; 5 cycles of 98 °C for 20 s, 65 °C for 20 s, 72 °C for 1 min; followed by 72 °C for 5 min and a 4 °C hold. The reaction was then split into two 12.5-µL aliquots. For the scRNA-seq pre-amplified library, the first aliquot was purified through two sequential SPRI clean-ups: first with 6.25 µL AMPure XP beads, then the supernatant was transferred to fresh 6.25 µL beads and purified again (each step included a 5-min room-temperature incubation, magnetic capture for 1 min, two washes with 200 µL 80% ethanol, 3-min air-drying, and elution in 22 µL water). The second aliquot, intended for scATAC-seq, was directly purified with 10 µL AMPure XP beads using the same washing and elution procedure, yielding the scATAC-seq pre-amplified library. Subsequently, 20 µL of each pre-amplified library was subjected to a final qPCR amplification with 20 µL NEBNext Master Mix, 0.5 µL P7-index primer, 0.5 µL of the respective indexing primer (N5-index for scATAC-seq or P5-solexa for scRNA-seq), and 0.4 µL 25× SYBR Green I, under the following cycling conditions: 98 °C for 2 min; cycles of 98 °C for 30 s, 63 °C for 30 s, 72 °C for 1 min. Both final libraries were purified with 40 µL AMPure XP beads (same wash/elution protocol) and eluted in 22 µL water. All libraries were quantified using a Qubit fluorometer (Q32851, ThermoFisher)and sequenced on an Illumina HiSeq X Ten platform with paired-end 150-bp reads.

736

737 **VEGFA stimulation and PDGFB expression analysis in HMC3 cells**

738 To perform VEGF induction experiments, HMC3 cells (or HEK293 cells as a non-microglial
739 control) were seeded in 6-well or 12-well plates one day prior to stimulation, using MEM
740 medium supplemented with 10% FBS and 40 ng/mL M-CSF. For VEGFA induction, the seeding
741 medium was replaced with MEM containing 10% FBS and 50 ng/mL recombinant VEGFA
742 (novoprotein, C744), while control cells received MEM with 1% BSA and 10% FBS. After 48 h
743 of stimulation, cells were harvested by trypsinization and total RNA was extracted using TRIzol.
744 Gene expression (specifically PDGFB) was quantified by real-time PCR using TB Green Premix
745 Ex Taq II (Tli RNase H Plus). Primers used in qPCR are listed in Supplementary Table 4.

746

747

748 **QUANTIFICATION AND STATISTICAL ANALYSIS**

749 **Parallel-split-seq data processing**

750 Parallel-split-seq data were processed using a customized pipeline encompassing read
751 preprocessing, barcode parsing, read mapping, and quality filtering. Briefly, cutadapt (v3.4)[127]
752 was utilized to trim raw reads and remove general sequencing adaptors. Cell barcodes were
753 parsed from Read1, allowing a single mismatch per barcode round as measured by the FREE
754 divergence algorithm[128]. For scRNA-seq libraries, reads containing mosaic end sequences
755 (indicating transposase contamination) were discarded. For scATAC-seq libraries, reads
756 possessing these sequences were retained. Reads were aligned to the human reference genome
757 (hg38) using STAR (v2.7.9a) [129], employing paired-end reads for scATAC-seq data and only
758 Read2 for scRNA-seq data.

759 Following mapping, a modified Python script derived from the SPLiT-seq[126] pipeline was
760 deployed to collapse unique molecular identifiers (UMIs) and quantify UMI counts for scRNA-
761 seq data, or unique fragment counts for scATAC-seq data, across all barcode combinations.
762 Reads mapping to the mitochondrial genome were removed from the scATAC-seq data prior to
763 further processing; for scRNA-seq data, mitochondrial reads were retained to enable quality
764 control.

765 Quality control was performed by quantifying detected genes in the scRNA-seq data and
766 calculating the transcription start site (TSS) enrichment score in the scATAC-seq data for each

unique barcode combination. For scRNA-seq, cells were excluded if they expressed fewer than 200 genes or if mitochondrial read counts exceeded 30% of total counts. For scATAC-seq, cells with a TSS enrichment score < 4 or with fewer than 500 unique fragments were discarded.

Dimension reduction, cell clustering, and annotation

The scRNA-seq count matrix was first library-size normalized to 10,000 counts per cell, log10-transformed, and used to identify highly variable genes. Principal component analysis (PCA) was performed on highly variable genes using the *scanpy.tl.pca* function with default parameters in Scanpy (v1.9.8)[130]. For scATAC-seq data, iterative latent semantic indexing (LSI) was performed using the *addIterativeLSI* function in ArchR (v1.0.2)[131] with parameters (2 iterations, LSI dimensions = 30). Batch effects, which were evident in the RNA PCA space but absent in ATAC features, were corrected by applying Harmony (implemented via *scanpy.external.pp.harmony_integrate*) to the RNA PCA embeddings with batch identity as the batch covariate. The top 30 Harmony-corrected principal components for RNA and the top 30 LSI dimensions for ATAC were subsequently integrated via weighted nearest neighbor (WNN) analysis using the *muon.pp.neighbors* function from muon (v0.1.5)[132]. Uniform Manifold Approximation and Projection (UMAP) was employed for visualization of the joint embedding. Cells were clustered on the joint WNN graph using the Leiden algorithm (*scanpy.tl.leiden*) on the joint graph. Cell types were assigned to each resulting cluster based on the expression of canonical marker genes.

Copy number variation (CNV) analysis

For scATAC-seq data, the genome was partitioned into 10-Mb sliding windows with a 2-Mb step size. Copy number alterations were estimated by quantifying Tn5 insertions within each window and comparing the signal against 100 GC content-matched background regions to infer copy number gains and losses[133]. For scRNA-seq data, CNVs were inferred using CopyKAT (v1.1.0)[134] with immune cells as the diploid reference population. PCA was performed independently on the RNA- and ATAC-derived CNV profiles, and the resulting principal components were integrated via WNN analysis to construct a joint CNV embedding. Leiden clustering was subsequently performed on the joint CNV graph to identify CNV-defined cell clusters. For each joint cluster, the mean absolute CNV signal was calculated separately for the

RNA and ATAC modalities and assigned to all constituent cells. The RNA- and ATAC-derived CNV scores were independently min-max normalized and then summed to generate a composite CNV score representing the overall level of aneuploidy for each cell.

Peak calling and genomic annotation for scATAC-seq data

Accessible chromatin peaks were identified from scATAC-seq data using MACS2 (v2.2.6)[135] through the *CallPeaks* function in the Signac (v.1.13.0)[60] with an effective genome size of 2.7×10^9 and a Tn5 offset correction (shift = -75 bp, extension = 150 bp). Peaks overlapping the ENCODE hg38 blacklist or located on non-standard chromosomes were excluded. A peak-by-cell count matrix was generated using the *FeatureMatrix* function in Signac and utilized for downstream analyses. Peaks were annotated using ChIPseeker (v1.38.0)[136] using the EnsDb.Hsapiens.v86 transcript annotation, defining promoter regions as −1 kbp to +100 bp relative to the TSSs. Putative CREs were further assigned by intersecting peaks with the GeneHancer Regulatory Elements Elite (hg38) annotation downloaded from the UCSC Genome Browser. Overlaps between peaks and GeneHancer promoter or enhancer annotations were identified using the *findOverlaps* function from the GenomicRanges (v1.54.1), requiring a minimum overlap of 100 bp.

Comparative atlas overlap analysis

To assess the reproducibility of scATAC-seq peaks across independent datasets, peak sets were compared against human cCREs from ENCODE database, the adult human brain chromatin accessibility cell atlas, and the fetal brain chromatin accessibility cell atlas. Reference peak sets were downloaded from their source repositories (https://screen.encodeproject.org/, https://decoder-genetics.wustl.edu/catlasv1/humanbrain/data/cCREs/cCREs.bed, https://catlas.org/catlasv1/catlas_downloads/humanbraindev/cCREs/consensus_peaks.bed). These bed files were subsequently converted into GRanges objects, and genomic overlaps between scATAC-seq peaks and each reference atlas were identified using GenomicRanges (v1.54.1).

Peak-gene cis-association and DORC identification

Peak-gene cis-associations were computed using FigR (v0.1.0)[68], which correlates chromatin accessibility with gene expression across single cells. Genes were ranked according to the number of significantly associated peaks ($P < 0.05$), and DORCs, defined as regulatory loci linked to genes with at least three significant peak associations, were subsequently identified. Before DORC score calculation, peak counts were normalized by the total number of unique mapping fragments within accessible peaks per cell. DORC scores were calculated as the summed accessibility of all significantly associated peaks linked to each gene, yielding a cell-by-gene DORC matrix.

Identification of differentially accessible chromatin regions

Differentially accessible chromatin regions (DACRs) were identified using the *FindMarkers* function in the Seurat (v.4.4.0) with logistic regression, incorporating the fraction of fragments in peaks as a latent variable to mitigate variations in sequencing depth across individual cells. *P* values were adjusted using Bonferroni correction across all evaluated peaks. Reference chromatin accessibility profiles of the fetal cerebrum and cerebellum were acquired from a human fetal chromatin accessibility atlas (Descartes), with genomic coordinates converted from hg19 to hg38 using the *liftOver* function in rtracklayer (v1.62.0). Cancer-specific DACRs were identified by comparing primary tumor cells against their most proximal normal lineages (astrocytes and oligodendrocytes for CLGG and DHGG; astrocytes for EPN; and cerebellar granule neurons for MB). The following parameters were strictly applied to the *FindMarkers* function: *min.pct = 0.01, min.diff.pct = 0, logfc.threshold = 0, only.pos = F*.

Quantification of super-enhancer overlap in DORC regions

To investigate the association between DORCs and super-enhancers (SEs), we performed a hypergeometric enrichment analysis for each DORC-associated gene. Overlap between the cCREs of DORC and curated SE regions was determined using the *findOverlaps* function in GenomicRanges (v1.54.1), ignoring strand information. cCREs overlapping SE regions were classified as SE-associated. For each gene, enrichment was evaluated based on the number of SE-associated cCREs among its DORC-linked cCREs. Specifically, if a gene was associated with n DORC cCREs, of which k overlapped SE regions, enrichment was assessed against the global DORC cCRE background (N total DORC cCREs, including K SE-associated cCREs)

using the hypergeometric distribution: *P (X⩾k) = phyper (k−1, K, N−K, n, lower.tail=FALSE)*. To account for multiple hypothesis testing across all genes, the resulting *P* values were adjusted using the Benjamini-Hochberg (BH) method to control the false discovery rate (FDR). Genes with $FDR < 0.05$ were considered significantly enriched for SE-associated DORC cCREs.

Transcription factor (TF) regulatory network and regulon scoring

TF regulators of DORCs were delineated using the *runFigRGRN* function from the FigR (v0.1.0). This method integrates two complementary metrics: enrichment of TF-binding motifs within DORC-associated peaks (based on the cisBP motif collection[68]) and the correlation between TF expressions and DORC scores across single cells. TFs with absolute regulation scores > 1 were retained as putative DORC regulators and used to construct a GRN.

Regulons, defined as individual TFs and their putative target genes inferred from TF–DORC associations in the GRN, were retained only if they contained at least four target genes. Regulon score was quantified using the AUCell algorithm implemented via pySCENIC (v0.12.1). Brefiely, AUCell quantifies regulon enrichment by calculating the area under the recovery curve (AUC) to assess whether the target genes of a regulon are preferentially enriched among the highest-ranked genes within the expression profile of each individual cell. The resulting matrix provides the AUC score for each regulon per cell. Cluster-specific regulons were further identified utilizing the regulon specificity score (RSS). For each regulon and cell type, the RSS was computed as $1 - \sqrt{JSD}$, where JSD represents the Jensen–Shannon divergence between the normalized distribution of regulon scores in cells of the target type and that in all other cells.

Cluster analysis and evaluation of tumor entity stratification

To compare the ability of different regulatory features to distinguish tumor entities, we performed unsupervised hierarchical clustering on per-patient profiles derived from three distinct feature matrices: aggregated regulon scores, average TF expressions, and ChromVAR-derived motif activity scores. For each feature type, pairwise Pearson correlation coefficients (r) between samples were calculated and converted into distances using the formula $d = (1 - r)/2$. Hierarchical clustering was then performed on the resulting distance matrices using the *Ward.D2* method implemented in the R *hclust* function. To facilitate comparison with the four known tumor entities represented in the cohort, dendrograms were cut into four clusters ($k = 4$).

To quantify the concordance between the clustering results and the cancer type annotations, we calculated the Adjusted Rand Index (ARI) using the R *mclust* package (v6.1.3). The ARI was computed by comparing the predicted cluster labels against the known cancer type labels for each sample.

Assignment of glioblastoma subtypes

Glioblastoma molecular subtypes (MES1-like, MES2-like, NPC1-like, NPC2-like, AC-like, and OPC-like) were assigned based on meta-modules previously defined[39] . For downstream analyses, MES1- and MES2-like cells were combined into a single MES-like group, whereas NPC1- and NPC2-like cells were collapsed into a single NPC-like group. Meta-module scores were calculated using scrabble (v1.0.0). Two-dimensional cellular state projections were generated, where each quadrant represents a distinct cellular state and exact coordinate positioning for individual cells was determined using hierarchical embedding implemented in scrabble.

Cell clustering based on regulon scores

To resolve cellular heterogeneity from the perspective of GRN activity, cells were clustered based on their regulon score profiles. The regulon score matrix (cells × regulons), generated from AUCell as described above, was used as the input feature space. A UMAP embedding was computed directly from this matrix using umap-learn (v0.5.5) with correlation distance as the similarity metric (*n_neighbors = 10* and *min_dist = 0.4*). Cells were subsequently partitioned into discrete regulon-defined clusters using the Leiden algorithm.

Pathway enrichment analysis

To identify Cancer Hallmark pathways enriched within each regulon, over-representation analysis was performed on the target genes of each regulon using the *gseapy.enrich* function from the Gseapy (v1.1.3).

Calculation of gene expression signature scores

Gene expression signature scores were computed using the *scanpy.tl.score_genes* function from Python package Scanpy (v1.9.8) with default parameters. Signature gene sets used for myeloid

cells were obtained from their respective original publications and are listed in Supplementary Table 5.

Survival analysis

RNA-seq expression profiles from the Children's Brain Tumor Tissue Consortium (CBTTC), including Low-grade glioma/astrocytoma (WHO grade I/II), High-grade glioma/astrocytoma (WHO grade III/IV), Medulloblastoma, and Ependymoma, were obtained from a published study[137]. Regulon activity was quantified for each bulk RNA-seq sample using AUCell (described above) based on the target gene sets of the regulons inferred from the single-cell data. Within each tumor type, patients were stratified into high- and low-regulon-activity groups according to the median regulon activity score. Progression-free survival and overall survival were evaluated using Kaplan–Meier analysis implemented in the survival (v3.5-8) and survminer (v0.4.9) R packages. Statistical significance was assessed using the log-rank test.

Somatic mutation calling from single-cell multi-omics data

To identify somatic mutations, BAM alignment files from both scRNA-seq and scATAC-seq were partitioned by patient prior to variant calling. Somatic variants were identified for each patient using SComatic, an algorithm designed for *de novo* mutation detection in single-cell data without a matched bulk normal, which systematically distinguishes true somatic mutations from sequencing artifacts and RNA editing events. High-confidence somatic variants were retained based on the SComatic filtering criteria.

Identification of mutation-gene associations in tumor cells

Mutation-gene association analysis was restricted to genomic sites for which both the reference and alternate alleles were detected in at least three tumor cells. Tumor cells were stratified by genotype, and differential gene expression between mutant and wild-type cells was assessed using *sc.tl.rank_genes_groups* function with *method='wilcoxon'*. To maximize association stringency, candidate genes were restricted to those where the mutation resided directly within their established cCREs. Genes with P value < 0.1 were deemed significantly associated with the corresponding mutation site.

SIREN analysis for microenvironmental interactions

Cell-cell interactions (CCIs) were inferred using the statistical analysis framework of CellPhoneDB (v4.0.0). Briefly, normalized single-cell RNA-seq count matrices and corresponding cell type annotations were utilized as inputs. The statistical significance of ligand-receptor interaction specificity across distinct cell populations was evaluated using a permutation test (iterations = 1,000), considering only genes expressed in at least 10% of cells within each cluster. Significant interactions were defined as those with $P < 0.05$. To delineate the microenvironmental signaling cascades targeting the primary cell population of interest, we specifically retained paracrine signaling events where ligands were expressed by non-tumor cells (lymphocytes, macrophages, oligodendrocytes, and microglia) and their cognate receptors were expressed on tumor cells.

To bridge extracellular signaling events with intracellular transcriptional reprogramming, tumor cell receptors were integrated with the gene regulatory network described above. Significant positive TF-target gene associations defined by FigR (regulation score > 1) were retained, and receptor-TF relationships were assigned using a curated receptor-TF reference database. This integration enabled the construction of a hierarchical regulatory axis mapping extracellular ligands to tumor receptors, and further to corresponding intracellular TFs and their downstream targets.

To empirically validate the functional coupling of the identified ligand-receptor-TF pathways within our specific biological context, the statistical correlation was evaluated between receptor expression and TF activity at single-cell resolution. Specifically, Pearson correlation coefficients were computed between the normalized gene expression matrix of the prioritized receptors and the continuous regulon activity scores of their linked TFs across all individual cells. *P* value derived from the Pearson correlation analysis were adjusted for multiple hypothesis testing using the Benjamini-Hochberg False Discovery Rate (FDR) procedure. Only Ligand-receptor-TF regulatory modules showing significant positive correlations ($FDR < 0.05$) were retained for downstream network modeling. Finally, the resulting ligand–receptor–TF–target relationships constituted the signal-to-regulon (SIREN) network.

Data availability

The raw FASTQ files of Parallel-seq data derived from human samples have been deposited in the Genome Sequence Archive (GSA) for Human at the National Genomics Data Center (NGDC), under project PRJCA068087.

Code availability

All the code used for computational prediction and data analysis is available at https://github.com/zhangqf-lab/SIREN.

Acknowledgements

This work has been supported by Tsinghua University Dushi Program (grant no. 20251080102 to Q.C.Z.), the Science & Technology Fundamental Resources Investigation Program (grant no. 2025FY100300 to Q.C.Z.), the National Natural Science Foundation of China (grants no. 32230018 and 32125007 to Q.C.Z., grants no. 81870834 and 62276027 to J.G.), the Shenzhen Science and Technology Program (grant no. AI2026009 to Q.C.Z.), the TsienTang Life Science Development Fund at Tsinghua University, the Beijing Advanced Innovation Center for Structural Biology, the XplorerPrize, and the Tsinghua-Peking Joint Center for Life Sciences. We thank the Postdoctoral Foundation of Tsinghua-Peking Center for Life Sciences (to Y.L.), and all members of the Q.C.Z labs their valuable discussions and contributions.

Author contributions

Q.C.Z and J.G. conceived the project, and Q.C.Z, J.G., and W.H. supervised the project. Y.L., and W.H. performed bioinformatics analyses, with the help of R.Y. and Y.S.. Y.W. and J.Z. conducted the Parallel-seq assays and validation experiments, with the help of K.F. and X.H.. L.T. contributed to the development and optimization of the Parallel-seq assay. Q.C.Z, Y.L., W.H., J.Z., and R.Y. wrote the manuscript with the input from all authors. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

References

1. Price, M. *et al.* CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2017-2021. *Neuro Oncol* **26**, vi1-vi85 (2024).
2. Louis, D.N. *et al.* The 2021 WHO Classification of Tumors of the Central Nervous System: a summary. *Neuro Oncol* **23**, 1231-1251 (2021).
3. Damodharan, S. & Puccetti, D. Pediatric Central Nervous System Tumor Overview and Emerging Treatment Considerations. *Brain Sci* **13**(2023).
4. Mackay, A. *et al.* Integrated Molecular Meta-Analysis of 1,000 Pediatric High-Grade and Diffuse Intrinsic Pontine Glioma. *Cancer Cell* **32**, 520-537.e5 (2017).
5. Ramaswamy, V. *et al.* Risk stratification of childhood medulloblastoma in the molecular era: the current consensus. *Acta Neuropathol* **131**, 821-31 (2016).
6. Monje, M. *et al.* Phase I trial of panobinostat in children with diffuse intrinsic pontine glioma: A report from the Pediatric Brain Tumor Consortium (PBTC-047). *Neuro Oncol* **25**, 2262-2272 (2023).
7. Stavinoha, P.L., Askins, M.A., Powell, S.K., Pillay Smiley, N. & Robert, R.S. Neurocognitive and Psychosocial Outcomes in Pediatric Brain Tumor Survivors. *Bioengineering (Basel)* **5**(2018).
8. Sleurs, C., Fletcher, P., Mallucci, C., Avula, S. & Ajithkumar, T. Neurocognitive Dysfunction After Treatment for Pediatric Brain Tumors: Subtype-Specific Findings and Proposal for Brain Network-Informed Evaluations. *Neurosci Bull* **39**, 1873-1886 (2023).
9. 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).
10. Neftel, C. *et al.* An Integrative Model of Cellular States, Plasticity, and Genetics for Glioblastoma. *Cell* **178**, 835-849 e21 (2019).
11. Hovestadt, V. *et al.* Resolving medulloblastoma cellular architecture by single-cell genomics. *Nature* **572**, 74-79 (2019).
12. Luo, Z.L. & Lu, R. Human Fetal Cerebellar Cell Atlas Informs Medulloblastoma Origin and Oncogenesis. *Neuro-Oncology* **25**(2023).
13. Smith, K.S. *et al.* Unified rhombic lip origins of group 3 and group 4 medulloblastoma. *Nature* **609**, 1012-1020 (2022).
14. Liu, I. *et al.* The landscape of tumor cell states and spatial organization in H3-K27M mutant diffuse midline glioma across age and location. *Nat Genet* **54**, 1881-1894 (2022).
15. Gillen, A.E. *et al.* Single-Cell RNA Sequencing of Childhood Ependymoma Reveals Neoplastic Cell Subpopulations That Impact Molecular Classification and Etiology. *Cell Rep* **32**, 108023 (2020).
16. Vladoiu, M.C. *et al.* Childhood cerebellar tumours mirror conserved fetal transcriptional programs. *Nature* **572**, 67-73 (2019).

17. Jessa, S. *et al.* Stalled developmental programs at the root of pediatric brain tumors. *Nature Genetics* **51**, 1702-1713 (2019).
18. Jessa, S. *et al.* K27M in canonical and noncanonical H3 variants occurs in distinct oligodendroglial cell lineages in brain midline gliomas. *Nat Genet* **54**, 1865-1880 (2022).
19. Northcott, P.A. *et al.* Enhancer hijacking activates GFI1 family oncogenes in medulloblastoma. *Nature* **511**, 428-34 (2014).
20. Mack, S.C. *et al.* Epigenomic alterations define lethal CIMP-positive ependymomas of infancy. *Nature* **506**, 445-50 (2014).
21. Schwartzentruber, J. *et al.* Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma. *Nature* **482**, 226-31 (2012).
22. Lewis, P.W. *et al.* Inhibition of PRC2 activity by a gain-of-function H3 mutation found in pediatric glioblastoma. *Science* **340**, 857-61 (2013).
23. Jones, D.T.W. *et al.* Molecular characteristics and therapeutic vulnerabilities across paediatric solid tumours. *Nat Rev Cancer* **19**, 420-438 (2019).
24. Gröbner, S.N. *et al.* The landscape of genomic alterations across childhood cancers. *Nature* **555**, 321-327 (2018).
25. Jones, C. *et al.* Pediatric high-grade glioma: biologically and clinically in need of new thinking. *Neuro Oncol* **19**, 153-161 (2017).
26. Filbin, M. & Monje, M. Developmental origins and emerging therapeutic opportunities for childhood cancer. *Nat Med* **25**, 367-376 (2019).
27. Zhao, S., Allis, C.D. & Wang, G.G. The language of chromatin modification in human cancers. *Nature Reviews Cancer* **21**, 413-430 (2021).
28. Timp, W. & Feinberg, A.P. Cancer as a dysregulated epigenome allowing cellular growth advantage at the expense of the host. *Nat Rev Cancer* **13**, 497-510 (2013).
29. Hanahan, D. Hallmarks of Cancer: New Dimensions. *Cancer Discov* **12**, 31-46 (2022).
30. Tirosh, I. & Suva, M.L. Cancer cell states: Lessons from ten years of single-cell RNA-sequencing of human tumors. *Cancer Cell* **42**, 1497-1506 (2024).
31. Grabovska, Y. *et al.* Pediatric pan-central nervous system tumor analysis of immune-cell infiltration identifies correlates of antitumor immunity. *Nature Communications* **11**, 4324 (2020).
32. DeCordova, S. *et al.* Molecular Heterogeneity and Immunosuppressive Microenvironment in Glioblastoma. *Front Immunol* **11**, 1402 (2020).
33. Jin, S., Plikus, M.V. & Nie, Q. CellChat for systematic analysis of cell–cell communication from single-cell transcriptomics. *Nature Protocols* **20**, 180-219 (2025).
34. Troulé, K. *et al.* CellPhoneDB v5: inferring cell–cell communication from single-cell multiomics data. *Nature Protocols* **20**, 3412-3440 (2025).
35. Browaeys, R., Saelens, W. & Saeys, Y. NicheNet: modeling intercellular communication by linking ligands to target genes. *Nature Methods* **17**, 159-162 (2020).
36. Ma, S. *et al.* Chromatin Potential Identified by Shared Single-Cell Profiling of RNA and Chromatin. *Cell* **183**, 1103-1116.e20 (2020).
37. Chen, X., Yang, W., Roberts, C.W.M. & Zhang, J. Developmental origins shape the paediatric cancer genome. *Nat Rev Cancer* **24**, 382-398 (2024).
38. Suvà, M.L. & Tirosh, I. The Glioma Stem Cell Model in the Era of Single-Cell Genomics. *Cancer Cell* **37**, 630-636 (2020).

39. Neftel, C. *et al.* An Integrative Model of Cellular States, Plasticity, and Genetics for Glioblastoma. *Cell* **178**, 835-849.e21 (2019).
40. Tang, L. *et al.* Joint analysis of chromatin accessibility and gene expression in the same single cells reveals cancer-specific regulatory programs. *Cell Syst* **16**, 101266 (2025).
41. Liu, L. *et al.* Deconvolution of single-cell multi-omics layers reveals regulatory heterogeneity. *Nature Communications* **10**, 470 (2019).
42. Cao, J. *et al.* Joint profiling of chromatin accessibility and gene expression in thousands of single cells. *Science* **361**, 1380-1385 (2018).
43. Zhu, C. *et al.* An ultra high-throughput method for single-cell joint analysis of open chromatin and transcriptome. *Nature Structural & Molecular Biology* **26**, 1063-1070 (2019).
44. Chen, S., Lake, B.B. & Zhang, K. High-throughput sequencing of the transcriptome and chromatin accessibility in the same cell. *Nature Biotechnology* **37**, 1452-1457 (2019).
45. Xu, W. *et al.* ISSAAC-seq enables sensitive and flexible multimodal profiling of chromatin accessibility and gene expression in single cells. *Nat Methods* **19**, 1243-1249 (2022).
46. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587.e29 (2021).
47. Korsunsky, I. *et al.* Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods* **16**, 1289-1296 (2019).
48. Gerdes, J. Ki-67 and other proliferation markers useful for immunohistological diagnostic and prognostic evaluations in human malignancies. *Semin Cancer Biol* **1**, 199-206 (1990).
49. Zhou, T., Niu, Y. & Li, Y. Advances in research on malignant tumors and targeted agents for TOP2A (Review). *Mol Med Rep* **31**(2025).
50. Oehlmann, M., Score, A.J. & Blow, J.J. The role of Cdc6 in ensuring complete genome licensing and S phase checkpoint activation. *J Cell Biol* **165**, 181-90 (2004).
51. Adini, A. *et al.* The stem cell marker prominin-1/CD133 interacts with vascular endothelial growth factor and potentiates its action. *Angiogenesis* **16**, 405-16 (2013).
52. Filbin, M.G. *et al.* Developmental and oncogenic programs in H3K27M gliomas dissected by single-cell RNA-seq. *Science* **360**, 331-335 (2018).
53. Bax, D.A. *et al.* EGFRvIII deletion mutations in pediatric high-grade glioma and response to targeted therapy in pediatric glioma cell lines. *Clin Cancer Res* **15**, 5753-61 (2009).
54. Mondal, G. *et al.* Pediatric bithalamic gliomas have a distinct epigenetic signature and frequent EGFR exon 20 insertions resulting in potential sensitivity to targeted kinase inhibition. *Acta Neuropathol* **139**, 1071-1088 (2020).
55. Boulay, G. *et al.* OTX2 Activity at Distal Regulatory Elements Shapes the Chromatin Landscape of Group 3 Medulloblastoma. *Cancer Discov* **7**, 288-301 (2017).
56. Cavalli, F.M.G. *et al.* Intertumoral Heterogeneity within Medulloblastoma Subgroups. *Cancer Cell* **31**, 737-754.e6 (2017).
57. Servidei, T. *et al.* Cell-of-Origin and Genetic, Epigenetic, and Microenvironmental Factors Contribute to the Intra-Tumoral Heterogeneity of Pediatric Intracranial Ependymoma. *Cancers (Basel)* **13**(2021).
58. Ahmadov, U. *et al.* A highly abundant circular RNA from the RMST locus plays a role in posterior fossa ependymoma pathogenesis. *Brain Pathology* **36**, e70071 (2026).

59. Feng, J., Liu, T., Qin, B., Zhang, Y. & Liu, X.S. Identifying ChIP-seq enrichment using MACS. *Nat Protoc* **7**, 1728-40 (2012).
60. Stuart, T., Srivastava, A., Madad, S., Lareau, C.A. & Satija, R. Single-cell chromatin state analysis with Signac. *Nature Methods* **18**, 1333-1341 (2021).
61. Heinz, S., Romanoski, C.E., Benner, C. & Glass, C.K. The selection and function of cell type-specific enhancers. *Nat Rev Mol Cell Biol* **16**, 144-54 (2015).
62. Harrow, J. *et al.* GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res* **22**, 1760-74 (2012).
63. Li, Y.E. *et al.* A comparative atlas of single-cell chromatin accessibility in the human brain. *Science* **382**, eadf7044 (2023).
64. Mannens, C.C.A. *et al.* Chromatin accessibility during human first-trimester neurodevelopment. *Nature* **647**, 179-186 (2025).
65. Terekhanova, N.V. *et al.* Epigenetic regulation during cancer transitions across 11 tumour types. *Nature* **623**, 432-441 (2023).
66. Alderton, G.K. Tumorigenesis: the origins of glioma. *Nat Rev Cancer* **11**, 627 (2011).
67. Taylor, M.D. *et al.* Radial glia cells are candidate stem cells of ependymoma. *Cancer Cell* **8**, 323-35 (2005).
68. Kartha, V.K. *et al.* Functional inference of gene regulation using single-cell multi-omics. *Cell Genom* **2**(2022).
69. Fishilevich, S. *et al.* GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. *Database (Oxford)* **2017**(2017).
70. Hanashima, C., Shen, L., Li, S.C. & Lai, E. Brain factor-1 controls the proliferation and differentiation of neocortical progenitor cells through independent mechanisms. *J Neurosci* **22**, 6526-36 (2002).
71. Ornitz, D.M. & Itoh, N. The Fibroblast Growth Factor signaling pathway. *Wiley Interdiscip Rev Dev Biol* **4**, 215-66 (2015).
72. Kiss, Z., Mudryj, M. & Ghosh, P.M. Non-circadian aspects of BHLHE40 cellular function in cancer. *Genes Cancer* **11**, 1-19 (2020).
73. Jin, L., Huang, S., Guan, C. & Chang, S. ETS1-activated SNHG10 exerts oncogenic functions in glioma via targeting miR-532-3p/FBXL19 axis. *Cancer Cell Int* **20**, 589 (2020).
74. Liang, J. *et al.* Ubiquitin-specific protease 22 promotes the proliferation, migration and invasion of glioma cells. *Cancer Biomark* **23**, 381-389 (2018).
75. Lötsch, D. *et al.* Targeting fibroblast growth factor receptors to combat aggressive ependymoma. *Acta Neuropathol* **142**, 339-360 (2021).
76. Adesina, A.M. *et al.* FOXG1 expression shows correlation with neuronal differentiation in cerebellar development, aggressive phenotype in medulloblastomas, and survival in a xenograft model of medulloblastoma. *Hum Pathol* **46**, 1859-71 (2015).
77. Lin, C.Y. *et al.* Active medulloblastoma enhancers reveal subgroup-specific cellular origins. *Nature* **530**, 57-62 (2016).
78. Mack, S.C. *et al.* Therapeutic targeting of ependymoma as informed by oncogenic enhancer profiling. *Nature* **553**, 101-105 (2018).
79. Sepp, M. *et al.* Cellular development and evolution of the mammalian cerebellum. *Nature* **625**, 788-796 (2024).

80. McDonough, A. *et al.* Unipolar (Dendritic) Brush Cells Are Morphologically Complex and Require Tbr2 for Differentiation and Migration. *Front Neurosci* **14**, 598548 (2020).
81. Tebbe, L. *et al.* Prph2 disease mutations lead to structural and functional defects in the RPE. *Faseb j* **36**, e22284 (2022).
82. Buenaventura, D.F., Corseri, A. & Emerson, M.M. Identification of Genes With Enriched Expression in Early Developing Mouse Cone Photoreceptors. *Invest Ophthalmol Vis Sci* **60**, 2787-2799 (2019).
83. Li, S., Qiu, F., Xu, A., Price, S.M. & Xiang, M. Barhl1 regulates migration and survival of cerebellar granule cells by controlling expression of the neurotrophin-3 gene. *J Neurosci* **24**, 3104-14 (2004).
84. Katsushima, K. *et al.* The long noncoding RNA lnc-HLX-2-7 is oncogenic in Group 3 medulloblastomas. *Neuro Oncol* **23**, 572-585 (2021).
85. Roussel, M.F. & Robinson, G.W. Role of MYC in Medulloblastoma. *Cold Spring Harb Perspect Med* **3**(2013).
86. Wang, D. & Owler, B.K. Expression of AQP1 and AQP4 in paediatric brain tumours. *J Clin Neurosci* **18**, 122-7 (2011).
87. Aibar, S. *et al.* SCENIC: single-cell regulatory network inference and clustering. *Nature Methods* **14**, 1083-1086 (2017).
88. Schep, A.N., Wu, B., Buenrostro, J.D. & Greenleaf, W.J. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat Methods* **14**, 975-978 (2017).
89. Owa, T. *et al.* Meis1 Coordinates Cerebellar Granule Cell Development by Regulating Pax6 Transcription, BMP Signaling and Atoh1 Degradation. *J Neurosci* **38**, 1277-1294 (2018).
90. Ayrault, O. *et al.* Atoh1 inhibits neuronal differentiation and collaborates with Gli1 to generate medulloblastoma-initiating cells. *Cancer Res* **70**, 5618-27 (2010).
91. Garancher, A. *et al.* NRL and CRX Define Photoreceptor Identity and Reveal Subgroup-Specific Dependencies in Medulloblastoma. *Cancer Cell* **33**, 435-449.e6 (2018).
92. Joshi, P. *et al.* Cell-state plasticity drives heterogeneity in Group 3/4 medulloblastoma. (bioRxiv, 2024).
93. Chen, A., Wang, H., Li, X., Anastasaki, C. & Gutmann, D.H. IRX2 and NPTX1 differential regulation of β-catenin underlies MEK-mediated proliferation in human neuroglial cells. *Genes Dev* **39**, 697-705 (2025).
94. Carro, M.S. *et al.* The transcriptional network for mesenchymal transformation of brain tumours. *Nature* **463**, 318-25 (2010).
95. Muzio, L. *et al.* Emx2 and Pax6 control regionalization of the pre-neuronogenic cortical primordium. *Cereb Cortex* **12**, 129-39 (2002).
96. Mikheeva, S.A. *et al.* TWIST1 promotes invasion through mesenchymal change in human glioblastoma. *Mol Cancer* **9**, 194 (2010).
97. Shapiro, J.A. *et al.* OpenPBTA: The Open Pediatric Brain Tumor Atlas. *Cell Genom* **3**, 100340 (2023).
98. Hornig, J. *et al.* The transcription factors Sox10 and Myrf define an essential regulatory network module in differentiating oligodendrocytes. *PLoS Genet* **9**, e1003907 (2013).

99. Schmidt-Edelkraut, U., Hoffmann, A., Daniel, G. & Spengler, D. Zac1 regulates astroglial differentiation of neural stem cells through Socs3. *Stem Cells* **31**, 1621-32 (2013).
100. Falcone, C., Filippis, C., Granzotto, M. & Mallamaci, A. Emx2 expression levels in NSCs modulate astrogenesis rates by regulating EgfR and Fgf9. *Glia* **63**, 412-22 (2015).
101. Lau, J. *et al.* STAT3 Blockade Inhibits Radiation-Induced Malignant Progression in Glioma. *Cancer Res* **75**, 4302-11 (2015).
102. Ijaz, H. *et al.* Pediatric high-grade glioma resources from the Children's Brain Tumor Tissue Consortium. *Neuro Oncol* **22**, 163-165 (2020).
103. Mandal, P., Belapurkar, V., Nair, D. & Ramanan, N. Vinculin-mediated axon growth requires interaction with actin but not talin in mouse neocortical neurons. *Cell Mol Life Sci* **78**, 5807-5826 (2021).
104. Leloup, L. & Wells, A. Calpains as potential anti-cancer targets. *Expert Opin Ther Targets* **15**, 309-23 (2011).
105. Huang, X. *et al.* CHL1 negatively regulates the proliferation and neuronal differentiation of neural progenitor cells through activation of the ERK1/2 MAPK pathway. *Mol Cell Neurosci* **46**, 296-307 (2011).
106. Byun, H. *et al.* Rorβ regulates selective axon-target innervation in the mammalian midbrain. *Development* **146**(2019).
107. Geering, B. Death-associated protein kinase 2: Regulator of apoptosis, autophagy and inflammation. *Int J Biochem Cell Biol* **65**, 151-4 (2015).
108. Saberiyan, M. *et al.* The role of DAPK2 as a key regulatory element in various human cancers: a systematic review. *Mol Biol Rep* **51**, 886 (2024).
109. Muyas, F. *et al.* De novo detection of somatic mutations in high-throughput single-cell profiling data sets. *Nature Biotechnology* **42**, 758-767 (2024).
110. Zhang, K.Y. *et al.* ASXL1 inactivation and reduced H3K27me3 across central nervous system tumors. *Acta Neuropathol* **148**, 19 (2024).
111. Wen, K., Zhu, W., Luo, Z. & Wang, W. Machine learning-based identification of histone deacetylase-associated prognostic factors and prognostic modeling for low-grade glioma. *Discov Oncol* **15**, 824 (2024).
112. Efremova, M., Vento-Tormo, M., Teichmann, S.A. & Vento-Tormo, R. CellPhoneDB: inferring cell–cell communication from combined expression of multi-subunit ligand–receptor complexes. *Nature Protocols* **15**, 1484-1506 (2020).
113. Zhang, Y. *et al.* CellCall: integrating paired ligand-receptor and transcription factor activities for cell-cell communication. *Nucleic Acids Res* **49**, 8520-8534 (2021).
114. Cheng, J., Zhang, J., Wu, Z. & Sun, X. Inferring microenvironmental regulation of gene expression from single-cell RNA sequencing data using scMLnet with an application to COVID-19. *Brief Bioinform* **22**, 988-1005 (2021).
115. Pitcher, J.L., Alexander, N., Miranda, P.J. & Johns, T.G. ErbB4 in the brain: Focus on high grade glioma. *Front Oncol* **12**, 983514 (2022).
116. Wang, C. *et al.* miR-124-3p and miR-194-5p regulation of the PI3K/AKT pathway via ROR2 in medulloblastoma progression. *Cancer Gene Ther* **31**, 941-954 (2024).
117. Shiraishi, R. *et al.* Cancer-specific epigenome identifies oncogenic hijacking by nuclear factor I family proteins for medulloblastoma progression. *Developmental Cell* **59**, 2302-2319.e12 (2024).

118. Masurkar, S.A., Deogharkar, A., Bharambe, H.S. & Shirsat, N.V. Downregulation of CRX, a Group 3-specific oncogenic transcription factor, inhibits TGF-β/activin signaling in medulloblastoma cells. *Biochem Biophys Res Commun* **568**, 76-82 (2021).
119. Sussman, J.H. *et al.* A longitudinal single-cell and spatial multiomic atlas of pediatric high-grade glioma. *Cell Rep Med* **7**, 102766 (2026).
120. Cheng, S. *et al.* A pan-cancer single-cell transcriptional atlas of tumor infiltrating myeloid cells. *Cell* **184**, 792-809.e23 (2021).
121. Whyte, W.A. *et al.* Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307-19 (2013).
122. Lovén, J. *et al.* Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* **153**, 320-34 (2013).
123. Bradner, J.E., Hnisz, D. & Young, R.A. Transcriptional Addiction in Cancer. *Cell* **168**, 629-643 (2017).
124. Barthel, F.P. *et al.* Systematic analysis of telomere length and somatic alterations in 31 cancer types. *Nature Genetics* **49**, 349-357 (2017).
125. Corces, M.R. *et al.* An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat Methods* **14**, 959-962 (2017).
126. Rosenberg, A.B. *et al.* Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. *Science* **360**, 176-182 (2018).
127. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10-12 (2011).
128. Hawkins, J.A., Jones, S.K., Jr., Finkelstein, I.J. & Press, W.H. Indel-correcting DNA barcodes for high-throughput sequencing. *Proc Natl Acad Sci U S A* **115**, E6217-e6226 (2018).
129. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).
130. Wolf, F.A., Angerer, P. & Theis, F.J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol* **19**, 15 (2018).
131. Granja, J.M. *et al.* ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat Genet* **53**, 403-411 (2021).
132. Bredikhin, D., Kats, I. & Stegle, O. MUON: multimodal omics analysis framework. *Genome Biol* **23**, 42 (2022).
133. Satpathy, A.T. *et al.* Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion. *Nat Biotechnol* **37**, 925-936 (2019).
134. Gao, R. *et al.* Delineating copy number and clonal substructure in human tumors from single-cell transcriptomes. *Nat Biotechnol* **39**, 599-608 (2021).
135. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137 (2008).
136. Yu, G., Wang, L.G. & He, Q.Y. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* **31**, 2382-3 (2015).
137. Zhang, Y., Chen, F., Donehower, L.A., Scheurer, M.E. & Creighton, C.J. A pediatric brain tumor atlas of genes deregulated by somatic genomic rearrangement. *Nat Commun* **12**, 937 (2021).

Fig. 1. Single-cell multi-omic profiling delineates cellular composition and regulatory landscape of pediatric brain tumors

(A) Schematic workflow of the study. Parallel single-cell profiling of chromatin accessibility and gene expression was performed across pediatric brain tumors. Integrative analyses were subsequently used to infer cancer-type-specific peak-to-gene associations, identify cis-regulatory element (CRE) mutations, reconstruct regulons, and map ligand-to-regulon cell-cell communication networks. Cohort composition includes medulloblastoma (MB, n = 13), circumscribed low-grade glioma (CLGG, n = 10), diffuse high-grade glioma (DHGG, n = 9), and ependymoma (EPN, n = 5).

(B) Heatmap summarizing the clinical and molecular information of the cohort, including sex, age, cancer type, cancer subtype, WHO grade, recurrence status, and survival outcome.

(C) WNN-UMAP visualization of 127,478 single cells colored by patient ID and grouped by cell lineages (tumor cells, oligodendrocytes, myeloid cells, and lymphocytes).

(D) WNN-UMAP showing the expression of representative marker genes (*MKI67*, *TOP2A*, *CDC6*, and *PROM1*) across the integrated embedding to highlight proliferative and stem-like tumor cell populations.

(E) Heatmap showing average marker gene expression across cell types and tumor categories.

(F) Pie charts showing the genomic distribution of candidate cis-regulatory elements (cCREs) identified in non-immune (Left) and immune (Right) cell populations.

(G) Bar plots illustrating the overlap between the cCREs identified in this study and established regulatory atlases (ENCODE, Human Adult Brain, and Human Fetal Brain).

(H) Pie chart depicting the genomic localization of differentially accessible cCREs (dacCREs).

(I) Bar plots illustrating the overall proportion of dacCREs in pediatric brain tumors and adult brain tumor (GBM).

Fig. 2. Landscape of cCRE-gene links and cancer-type-specific regulatory programs

(A) Circos plot (Left) illustrating the global landscape of chromatin accessibility and gene regulatory links of MB across the genome. Tracks from outer to inner represent: chromosomes, DNase-seq (D721Med cell line reference), scATAC-seq signals, identified peak-gene associations, and scRNA-seq expression levels. Representative oncogenic drivers and TFs (e.g.,

MYCN, PAX6, BARHL1) are highlighted. An enlarged view (Right) of chromosome 9 showcases high-resolution mapping of peak-gene associations (blue lines) linked to target genes such as BARHL1. A bar plot (Middle top) shows the number of genes associated with different numbers of cCREs.

(B) Bar plot showing the distribution of cCRE-gene links across cancer types (DHGG, CLGG, EPN, and MB), categorized into promoter regions, GeneHancer-annotated enhancers, and previously unannotated regions.

(C) UpSet plot showing the intersections of cCRE-gene links across the four cancer types.

(D) Scatter plots correlating differential cCRE chromatin accessibility with differential expression of linked genes. Each dot represents a cCRE-gene pair; the x-axis and y-axis indicate log2(fold change) in gene expression and scATAC-seq peak accessibility (tumor versus normal cells), respectively. Red lines indicate linear regression fits; gray shading represents 95% confidence intervals.

(E-F) Dot plots showing chromatin accessibility of dacCREs linked to representative genes, highlighting cancer-acquired accessibility gains (E) and losses (F) relative to putative developmental origins. DacCREs are grouped by specificity. Dot size indicates the fraction of cells with accessible peaks. Color intensity denotes log2(fold change) in chromatin accessibility.

Fig. 3. DORCs reveal the regulatory landscape across pediatric brain tumor types.

(A) Schematic workflow defining DORCs and their overlap with super-enhancers (SEs). DORCs are defined as genomic regions associated with genes linked to more than three significant cCREs. Overlap significance of DORCs with SEs is determined by hypergeometric test.

(B-C) Scatter plot shows the enrichment and significance of DORCs overlapped known cancer-specific SEs in MB (B) and EPN (C). Genes associated with DORCs overlapping previously curated cancer-specific SEs are highlighted. Dot size represents the number of associated cCREs; color intensity shows fold enrichment.

(D) Venn diagrams showing the overlap of DORCs (pink) and SEs (blue) across cancer types. Question marks indicate that corresponding SE annotations were unavailable for CLGG and DHGG.

(E–I) Representative track plots showing regulatory landscapes for subtype-specific regulators: BARHL1 in MB (E), HLX in MB (F), MYCN in MB (G), CACNA1H in EPN (H), and AQP1 in

EPN (I). Left panels show normalized scATAC-seq signals, reference DNase-seq, and SE annotations. Right panels display violin plots comparing DORC activity scores and gene expression levels across subtypes (red highlights indicate the specifically activated subtype).

Fig. 4. Regulon activity resolves cancer types and stratifies clinical outcomes in pediatric brain tumors.

(A) UMAP embeddings of MB tumor cells based on regulon scores, colored by patient ID (Left) and MB subtype (Right).

(B) Heatmap shows unsupervised hierarchical clustering of 36 patients (HGG02 excluded) based on aggregate regulon scores. Representative lineage-specific regulons are highlighted on the right (SHH-specific, green; Group 3-specific, red; Group 4-specific, brown).

(C) Ranking of regulons by regulon specificity scores (RSS) for MB subtypes. Top-10-ranked specific regulons are labeled, with key master regulators discussed in the main text highlighted in red.

(D) UMAP embedding of DHGG cells. Unsupervised Leiden clustering on regulon scores reveals 10 discrete clusters (Left). Cells are colored by predicted transcriptomic cell states (Right) based on established glioblastoma meta-modules (AC-like, MES-like, NPC-like, and OPC-like).

(E) Ranking of regulons by RSS for each DHGG cell state. Top-10-ranked, state-specific regulons are labeled, with key master regulators discussed in the main text highlighted in red.

(F) Volcano plot displaying the hazard ratio (HR) against the -log10 (*P* value) derived from survival analysis of individual regulon in MB. Light gray dots denote non-significant regulons. Blue dots indicate regulons with high regulon scores are associated with favorable survival (HR < 1), while red dots indicate regulons with low regulon scores are associated with favorable survival (HR > 1). Key prognostic regulons (including MYC and PAX6) are annotated. The dashed lines indicate the thresholds for significance and an HR of 1.

(G) Survival analysis stratified by MYC and PAX6 regulon activity. Kaplan-Meier survival curves show overall survival probability over time (months) for MB patients stratified by high (red line) versus low (blue line) regulon activity of MYC (Top) and PAX6 (Bottom). Statistical significance was determined using the log-rank test, with *P* value indicated in each panel.

(H) Network diagrams illustrating target genes regulated by MYC (Top) and PAX6 (Bottom). Red squares represent the transcription factors (TFs), and blue squares represent their respective

target genes. Green connecting lines (edges) indicate inferred regulatory links, with line thickness corresponding to regulon scores.

Fig. 5. Somatic mutations are enriched in cis-regulatory elements and are associated with target gene expression

(A) Bar plot illustrating the number of somatic mutations identified from scATAC-seq data across 25 patients in different cell types (tumor cell, macrophage, microglia, lymphocyte, and oligodendrocyte).

(B-C) Comparison of mutation characteristics between tumor and non-tumor cells. Violin plots comparing percentages of deleterious mutations (B) and conservation (phyloP7way) scores (C) between tumor and non-tumor cells.

(D) Violin plot displaying the fold enrichment of mutation density within cCREs linked to genes relative to unlinked cCREs across the four cancer types.

(E) Manhattan Plot showing the significance of differentially expressed genes with cCRE mutations. Significant candidates are highlighted. Blue and red dots denote significantly upregulated and downregulated genes, respectively, in cCRE-mutated versus wild-type cells, respectively. Horizontal dashed line indicates the significance threshold (*P* value = 0.1).

(F-G) Representative track plots of eCRE-mutations in DHGG (F) and MB (G). Left panels show genomic scATAC-seq signals, identified peaks, and peak-to-gene regulatory links with specific mutations indicated. Right panels display violin plots quantifying target gene expression in mutated (red) versus wild-type (blue) cells.

Fig. 6. Integration of cell-cell communication and regulon activity reveals signaling-to-transcription networks

(A) Schematic representation of the signaling-to-regulon inference framework. The model illustrates a three-layer regulatory cascade: (1) Sender cells (Left) express specific ligands; (2) these ligands interact with cognate receptors on the receiver cells (Right); (3) receptor activation triggers an intracellular signal transduction that modulates the activity of downstream TF regulons.

(B) Dot plot (Left) showing the inferred ligand-receptor (L-R) interactions between DHGG OPC-like cells and various microenvironment cell types. Color intensity and dot size represents

the interaction strength. Red circles indicate statistical significance. Heatmaps display the correlations between TF regulon activity and receptor expression (Right top) or signaling-related pathways (Right bottom).

(C) A comprehensive signaling-to-regulon network for DHGG illustrating information flow from extracellular ligands (green squares), to cognate receptors (brown squares), intracellular master TFs (orange squares), and downstream target genes (blue squares).

(D) Dot plot (Top) showing the inferred ligand-receptor (L-R) interactions between microglia subpopulations and distinct glioblastoma cellular states (including AC-like, MES-like, NPC-like, and OPC-like cells). Color intensity and dot size represents the interaction strength. Red circles indicate statistical significance. Heatmaps (Bottom) display the correlations between TF regulon activity and receptor expression.

(E) A comprehensive signaling-to-regulon network for DHGG microglia illustrating information flow from extracellular ligands (green squares), to cognate receptors (brown squares), intracellular master TFs (orange squares), and downstream target genes (blue squares).

(F) Experimental validation of VEGFA-mediated induction of PDGFB. Bar plot showing the fold change of PDGFB expression determined by RT-qPCR in HMC3 (human microglia cell line) and HEK293 (control) cells treated with either 1% BSA (vehicle control) or 50 ng/mL recombinant human VEGFA. Statistical significance was evaluated using two-tailed Student's *t*-test, *** $p < 0.001$, n.s., not significant.

Extended Data Fig. 1. Quality control of joint scRNA and scATAC profiling and characterization of cell lineages.

(A) Violin plots showing the distribution of RNA-based quality metrics across cancer types, including total UMI counts, number of detected genes, and percentage of mitochondrial reads per cell.

(B) Violin plots showing the distribution of ATAC-based quality metrics across cancer types, including the number of fragments, transcription start site (TSS) enrichment scores, and fraction of reads in peaks.

(C) Aggregate Tn5 insertion profiles centered on TSSs across cancer types.

(D) Distribution of ATAC-seq fragment sizes across cancer types.

(E) WNN-UMAP visualization of all single cells colored by cancer type, illustrating multi-omic data integration.

(F) WNN-UMAP feature plots showing the expression of representative cell type marker genes, including oligodendrocyte (*OPALIN*), myeloid (*CD68*), T cell (*CD3D*), B cell (*CD79A*), and cancer-associated genes (*OTX2*, *RMST*, *AQP4*, *PDGFRA*, *EGFR*, *MYCN*).

(G) WNN-UMAP visualization of inferred copy number variation (CNV) scores, distinguishing malignant from non-malignant cell populations.

(H) WNN-UMAP visualization of single cells colored by sample identity (Left) and inferred CNV scores (Right).

(I) Pie chart depicting the genomic localization of dacCREs across four pediatric brain tumor types and adult brain tumor (GBM).

(J) Heatmaps displaying the log2(fold change) of scATAC-seq peak accessibility in tumor relative to normal cells across cancer types.

(K) Bar plots illustrating Hallmark pathway enrichment separately for pediatric DHGG-specific and adult GBM-specific regulatory gene sets.

Extended Data Fig. 2. Landscape of cCRE–gene associations and cancer-type-specific regulatory programs across cancer types.

(A-C) Circos plots illustrating global landscape of chromatin accessibility and gene regulatory links of CLGG (A), DHGG (B), and EPN (C). Tracks (outer to inner) represent: chromosomes, scATAC-seq signal (orange), identified peak-gene associations (blue lines), scRNA-seq expression levels (purple), and gene annotations (green). Key lineage-associated genes and regulatory TFs (e.g., BHLHE40 in CLGG, MEIS2 in DHGG, and CACNA1H in EPN) are labeled. A bar plot (center) shows the number of genes with distinct correlated associations.

(D-E) Dot plots showing differential expression of representative genes linked to dacCREs (related to Figure 2E-F). Dot size indicates the fraction of expressing cells; color intensity represents the log2 fold change in gene expression between tumor and reference cells.

(F-G) Representative track plots displaying chromatin accessibility profiles, dacCREs, and co-accessibility links for the FGFR1 (F) and FOXG1 (G) loci. Normalized scATAC-seq tracks (Left) and single-cell gene expression violin plots (Right) are stratified across the four tumor entities and normal reference lineages. Shaded vertical bars highlight lineage-specific dacCREs (orange

blocks). "Links" panels depict predicted cis-regulatory interactions connecting dacCREs to target gene promoters. Gene models and genomic coordinates (Mb) are shown at the bottom.

Extended Data Fig. 3. Identification of DORCs across pediatric brain tumor subtypes.
Genes are ranked by the number of significant cCRE-gene links (correlated peaks). Genomic regions linked to genes with more than 3 associated cCREs are defined as DORCs. Top-ranked DORC-related genes are highlighted in red.

Extended Data Fig. 4. Single-cell regulon profiles resolve intratumoral heterogeneity and subtype-specific drivers.
(A-B) Violin plots summarizing the global architecture of the gene regulatory network (GRN) across four cancer types, showing the distribution of unique TF regulators per target gene (A) and target genes per TF (B).
(C-E) UMAP embeddings of CLGG (C), DHGG (D), and EPN (E) single cells based on regulon scores, colored by patient ID or cancer subtype.
(F-G) Comparison of patient stratification across different multi-omic features, corresponding to Figure 4B. Heatmaps show gene expression profiles (F) and inferred TF activities based on ChromVAR motif accessibility (G) across 36 patients.
(H-I) Unsupervised hierarchical clustering of 36 patients based on gene expression (H) and ChromVAR scores (I).
(J-K) Ranking of regulons by RSS for the corresponding subtypes in DHGG (J) and EPN (K). Top-10-ranked, subtype-specific regulons are labeled, with key master regulators discussed in the main text highlighted in red.
(L) UMAP embeddings of MB single cells based on regulon scores, colored by cancer subtype.
(M) Ranking of regulons by RSS for the corresponding subtypes in MB Group 3 (G3a, G3b). Top-10-ranked, subtype-specific regulons are labeled in red.
(N) UMAP embeddings of an independent MB clinical cohort based on regulon scores, colored by clinically annotated cancer type (left), cancer subtype assigned by correlation with reference profiles from our single-cell dataset (right).

(O) Kaplan–Meier curves showing overall survival of Group 3 MB patients stratified by the regulon-based G3a and G3b regulatory subtypes. Survival differences were assessed using the log-rank test.

Extended Data Fig. 5. Regulon activity resolves intratumoral heterogeneity and defines TF-driven cell states in DHGG.

(A) 2D star plots showing the gene signature score for each cell. The four corners of the plots represent the four cellular identities: Astrocyte-like (AC-like), Mesenchymal-like (MES-like), Oligodendrocyte precursor-like (OPC-like), and Neural-progenitor-like (NPC-like). Color intensity represents the gene signature score for each specific state.

(B) UMAP embedding plots showing the regulon activity of the four cell states (AC-like, MES-like, NPC-like, and OPC-like).

(C) Heatmap showing the proportion of cells from each Leiden cluster assigned to the four cell states.

(D) Ranking of regulons by RSS for the 10 unsupervised Leiden clusters (Clusters 0-9) identified in DHGG. Top-10-ranked, subtype-specific regulons are highlighted in red.

Extended Data Fig. 6. Identification of subtype-specific prognostic regulons and their impact on patient survival.

(A) Volcano plot displaying the hazard ratio (HR) against the -log10(*P* value) derived from survival analysis of individual regulon in CLGG, DHGG, and EPN. Each dot represents a regulon. Light gray dots denote non-significant regulons. Blue dots indicate regulons with high regulon scores are favorable (HR < 1), while red dots indicate regulons with low regulon scores are favorable (HR > 1). Key prognostic regulons are annotated. The dashed lines indicate the thresholds for significance and an HR of 1.

(B) Survival analysis stratified by FOXG1 (CLGG), BCL11B (DHGG), EMX2 (DHGG), and TEAD2 (EPN) regulon activity. Kaplan-Meier survival curves show overall survival probability over time (months) for patients stratified by high (red line) versus low (blue line) regulon activity of four regulons. Statistical significance was determined using the log-rank test, with *P* value indicated in each panel.

Extended Data Fig. 7. Somatic mutational landscape across cellular lineages.

(A) Bar plot illustrating the number of somatic mutations identified from scRNA-seq data across 25 patients in different cell types (tumor cell, macrophage, microglia, lymphocyte, and oligodendrocyte).

(B) Bar plot illustrating the number of mutations categorized by their localization within identified cCREs (red) or other chromatin-accessible regions (light blue) for each patient.

(C) Bar plot illustrating the genomic distribution of mutations in tumor versus non-tumor cells detected from scRNA-seq data.

(D) Bar plot illustrating the percentage of mutations predicted as deleterious or tolerant across samples.

(E) Violin plots showing the distribution of sequencing (scATAC-seq) counts between gene-linked (red) and non-gene-linked (blue) cCREs.

(F) Gene Ontology (GO) enrichment analysis of genes associated with cCRE-mutations for four cancer types.

Extended Data Fig. 8. Characterization of immune cell diversity and the signaling-to-regulon framework.

(A) UMAP embedding of immune cells, colored by high-resolution cell type annotations.

(B) Dot plot showing the expression of canonical marker genes used to define the immune sub-populations.

(C-F) WNN-UMAP feature plots illustrating the scoring of functional modules and specific TF activity within the myeloid compartment, including pro-inflammatory signature scores (C), *ex vivo* activation versus homeostasis signatures (D), NFKB1 regulon activity (E), and specific functional scores for M1/M2 polarization, angiogenesis, and phagocytosis (F).

Extended Data Fig. 9. Integration of cell-cell communication and regulon activity reveals signaling-to-transcription networks in MB.

(A) Dot plot (Top) showing the inferred ligand-receptor (L-R) interactions between malignant SHH MB cells and various microenvironment cell types. Color intensity and dot size represents the interaction strength. Red circles indicate statistical significance. Heatmaps display the

correlation between TF regulon activity and receptor expression (Middle) or signaling-related pathways (Bottom).

(B) A comprehensive signaling-to-regulon network for MB illustrating information flow from extracellular ligands (green squares), to cognate receptors (brown squares), intracellular master TFs (orange squares), and downstream target genes (blue squares).

(C) Dot plot (Left) showing the inferred ligand-receptor (L-R) interactions between malignant MB Group 3 cells and various microenvironment cell types. Color intensity and dot size represents the interaction strength. Red circles indicate statistical significance. Heatmaps display the correlation between TF regulon activity and receptor expression (Right top) or signaling-related pathways (Right bottom).

Extended Data Fig. 10. Subtype-specific signaling cascades and regulon coordination.

(A-B) Dot plot (Left) showing the inferred ligand-receptor (L-R) interactions between various microenvironment cell types and malignant DHGG AC-like (A) and DHGG MES-like (B) cells. Color intensity and dot size represents the interaction strength. Red circles indicate statistical significance. Heatmaps display the correlation between TF regulon activity and receptor expression (Right top) or signaling-related pathways (Right bottom).

Fig. 1

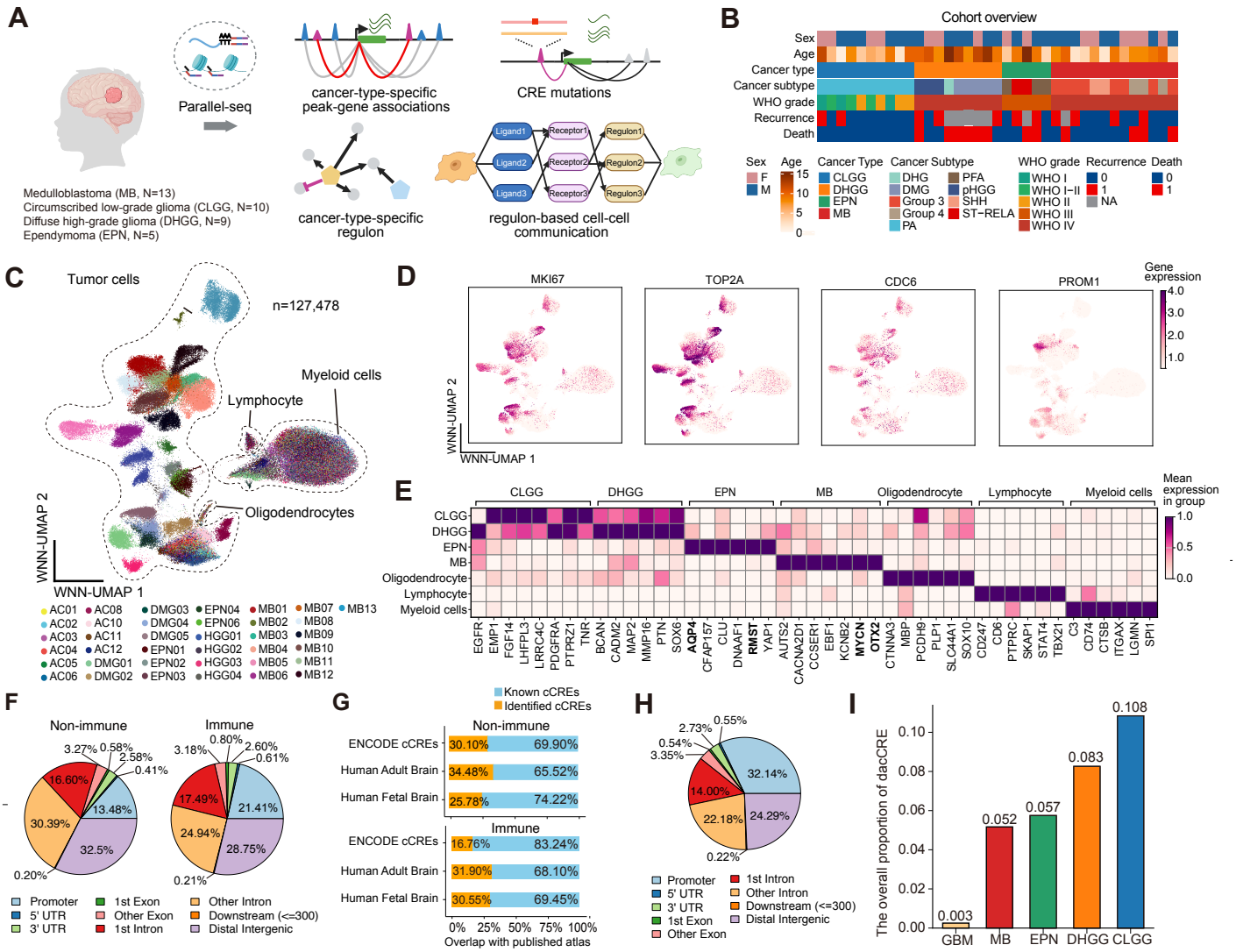

A
Parallel-seq
cancer-type-specific peak-gene associations
CRE mutations
cancer-type-specific regulon
regulon-based cell-cell communication
Ligand1
Ligand2
Ligand3
Receptor1
Receptor2
Receptor3
Regulon1
Regulon2
Regulon3
Medulloblastoma (MB, N=13)
Circumscribed low-grade glioma (CLGG, N=10)
Diffuse high-grade glioma (DHGG, N=9)
Ependymoma (EPN, N=5)
B
Cohort overview
Sex
Age
Cancer type
Cancer subtype
WHO grade
Recurrence
Death
F
M
CLGG
DHGG
EPN
MB
DHG
DMG
Group 3
Group 4
PA
PFA
pHGG
SHH
ST−RELA
WHO I
WHO I−II
WHO II
WHO III
WHO IV
NA
C
Tumor cells
n=127,478
Myeloid cells
Lymphocyte
Oligodendrocytes
WNN-UMAP 1
WNN-UMAP 2
AC01 AC02 AC03 AC04 AC05 AC06 AC08 AC10 AC11 AC12 DMG01 DMG02 DMG03 DMG04 DMG05 EPN01 EPN02 EPN03 EPN04 EPN06 HGG01 HGG02 HGG03 HGG04 MB01 MB02 MB03 MB04 MB05 MB06 MB07 MB08 MB09 MB10 MB11 MB12 MB13
D
MKI67
TOP2A
CDC6
PROM1
Gene expression
E
CLGG
DHGG
EPN
MB
Oligodendrocyte
Lymphocyte
Myeloid cells
Mean expression in group
EGFR EMP1 FGF14 LHFPL3 LRRC4C PDGFRA PTPRZ1 TNR BCAN CADM2 MAP2 MMP16 PTN SOX6 AQP4 CFAP157 CLU DNAAF1 RMST YAP1 AUTS2 CACNA2D1 CCSER1 EBF1 KCNB2 MYCN OTX2 CTNNA3 MBP PCDH9 PLP1 SLC44A1 SOX10 CD247 CD6 PTPRC SKAP1 STAT4 TBX21 C3 CD74 CTSB ITGAX LGMN SPI1
F
Non-immune
Immune
3.27% 0.58% 2.58% 0.41% 16.60% 13.48% 30.39% 32.5% 0.20%
3.18% 0.80% 2.60% 0.61% 17.49% 21.41% 24.94% 28.75% 0.21%
Promoter
5' UTR
3' UTR
1st Exon
Other Exon
1st Intron
Other Intron
Downstream (<=300)
Distal Intergenic
G
Known cCREs
Identified cCREs
Non-immune
ENCODE cCREs 30.10% 69.90%
Human Adult Brain 34.48% 65.52%
Human Fetal Brain 25.78% 74.22%
Immune
ENCODE cCREs 16.76% 83.24%
Human Adult Brain 31.90% 68.10%
Human Fetal Brain 30.55% 69.45%
0% 25% 50% 75% 100%
Overlap with published atlas
H
2.73% 0.55% 0.54% 3.35% 32.14% 14.00% 22.18% 24.29% 0.22%
Promoter
5' UTR
3' UTR
1st Exon
Other Exon
1st Intron
Other Intron
Downstream (<=300)
Distal Intergenic
I
The overall proportion of dacCRE
0.003 0.052 0.057 0.083 0.108
GBM MB EPN DHGG CLGG

Fig. 2

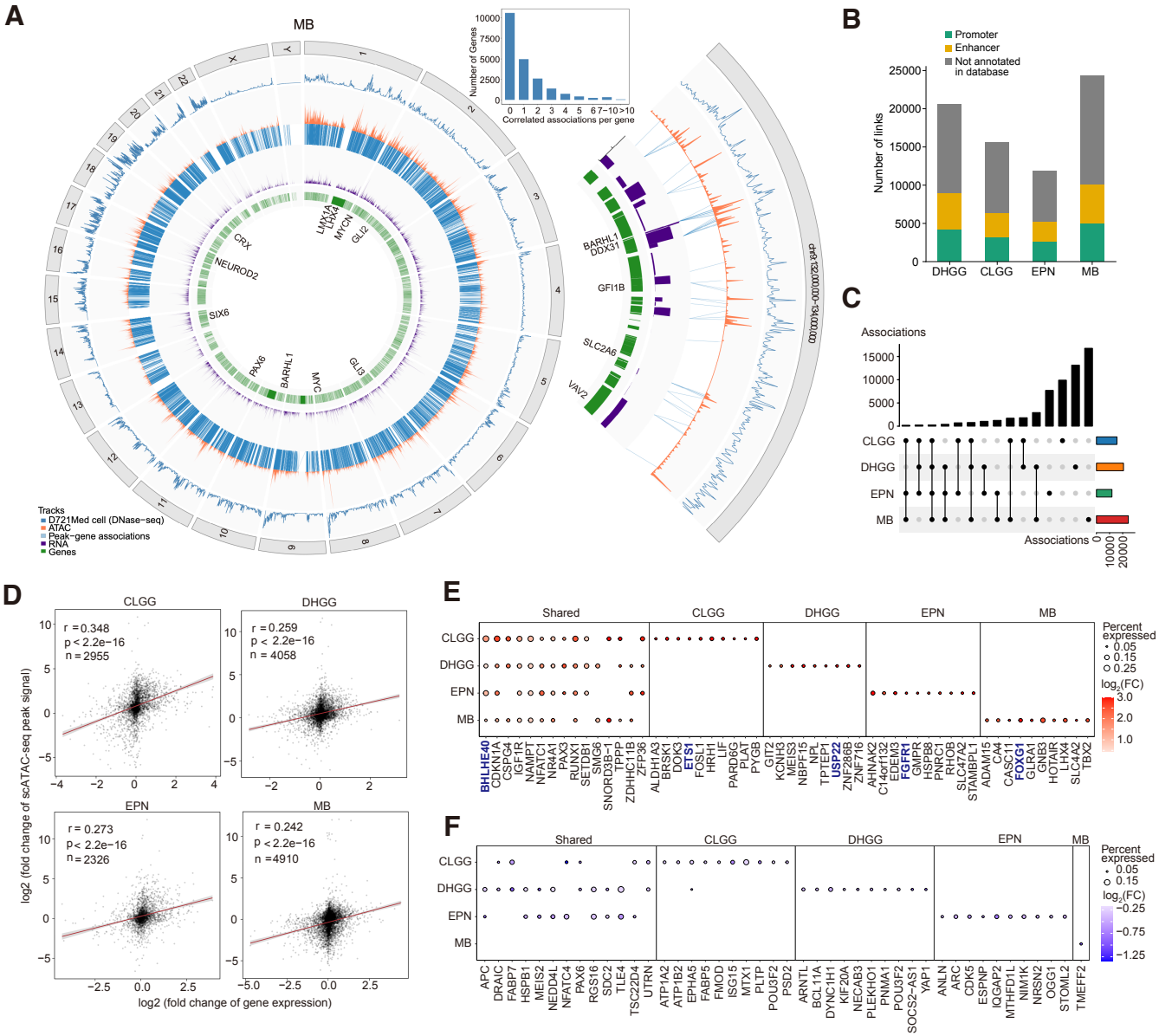

A
MB
Number of Genes
Correlated associations per gene
Tracks
D721Med cell (DNase−seq)
ATAC
Peak−gene associations
RNA
Genes
B
Promoter
Enhancer
Not annotated in database
Number of links
DHGG
CLGG
EPN
MB
C
Associations
Associations
D
CLGG
DHGG
EPN
MB
log2 (fold change of scATAC-seq peak signal)
log2 (fold change of gene expression)
E
Shared
CLGG
DHGG
EPN
MB
Percent expressed
log2(FC)
F
Shared
CLGG
DHGG
EPN
MB
Percent expressed
log2(FC)

Fig. 3

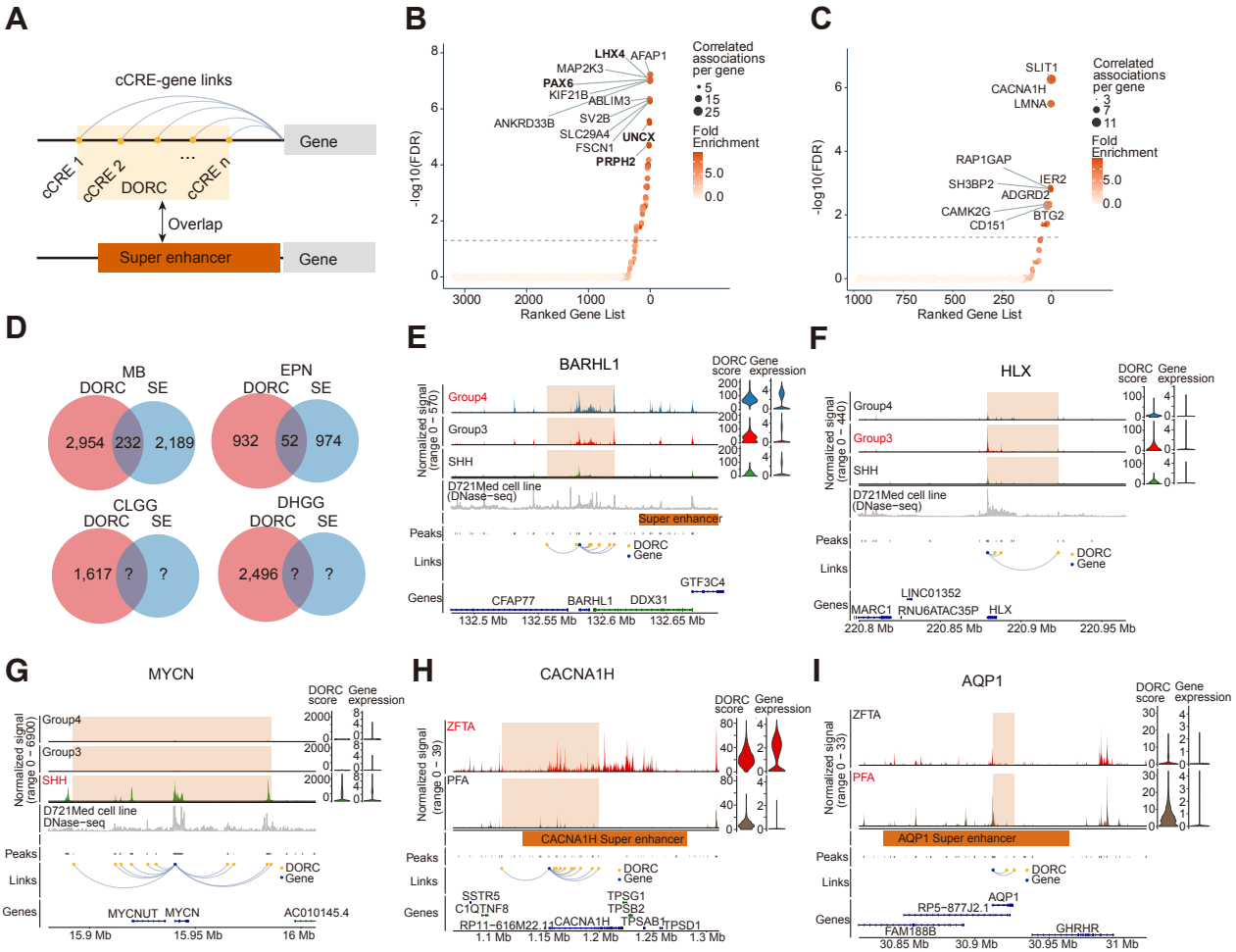

A
cCRE-gene links
Gene
cCRE 1
cCRE 2
DORC
cCRE n
Overlap
Super enhancer
Gene
B
LHX4
AFAP1
MAP2K3
PAX6
KIF21B
ABLIM3
ANKRD33B
SV2B
SLC29A4
UNCX
FSCN1
PRPH2
Correlated associations per gene
5
15
25
Fold Enrichment
5.0
0.0
-log10(FDR)
Ranked Gene List
C
SLIT1
CACNA1H
LMNA
RAP1GAP
SH3BP2
IER2
ADGRD2
CAMK2G
CD151
BTG2
Correlated associations per gene
3
7
11
Fold Enrichment
5.0
0.0
-log10(FDR)
Ranked Gene List
D
MB
DORC
SE
2,954
232
2,189
EPN
DORC
SE
932
52
974
CLGG
DORC
SE
1,617
?
?
DHGG
DORC
SE
2,496
?
?
E
BARHL1
DORC score
Gene expression
Normalized signal (range 0 – 570)
Group4
Group3
SHH
D721Med cell line (DNase−seq)
Super enhancer
Peaks
Links
DORC
Gene
Genes
GTF3C4
CFAP77
BARHL1
DDX31
132.5 Mb
132.55 Mb
132.6 Mb
132.65 Mb
F
HLX
DORC score
Gene expression
Normalized signal (range 0 – 440)
Group4
Group3
SHH
D721Med cell line (DNase−seq)
Peaks
Links
DORC
Gene
Genes
LINC01352
MARC1
RNU6ATAC35P
HLX
220.8 Mb
220.85 Mb
220.9 Mb
220.95 Mb
G
MYCN
DORC score
Gene expression
Normalized signal (range 0 – 6900)
Group4
Group3
SHH
D721Med cell line DNase−seq
Peaks
Links
DORC
Gene
Genes
MYCNUT
MYCN
AC010145.4
15.9 Mb
15.95 Mb
16 Mb
H
CACNA1H
DORC score
Gene expression
Normalized signal (range 0 – 39)
ZFTA
PFA
CACNA1H Super enhancer
Peaks
Links
DORC
Gene
Genes
SSTR5
C1QTNF8
RP11−616M22.1
CACNA1H
TPSG1
TPSB2
TPSAB1
TPSD1
1.1 Mb
1.15 Mb
1.2 Mb
1.25 Mb
1.3 Mb
I
AQP1
DORC score
Gene expression
Normalized signal (range 0 – 33)
ZFTA
PFA
AQP1 Super enhancer
Peaks
Links
DORC
Gene
Genes
AQP1
RP5−877J2.1
FAM188B
GHRHR
30.85 Mb
30.9 Mb
30.95 Mb
31 Mb

Fig. 4

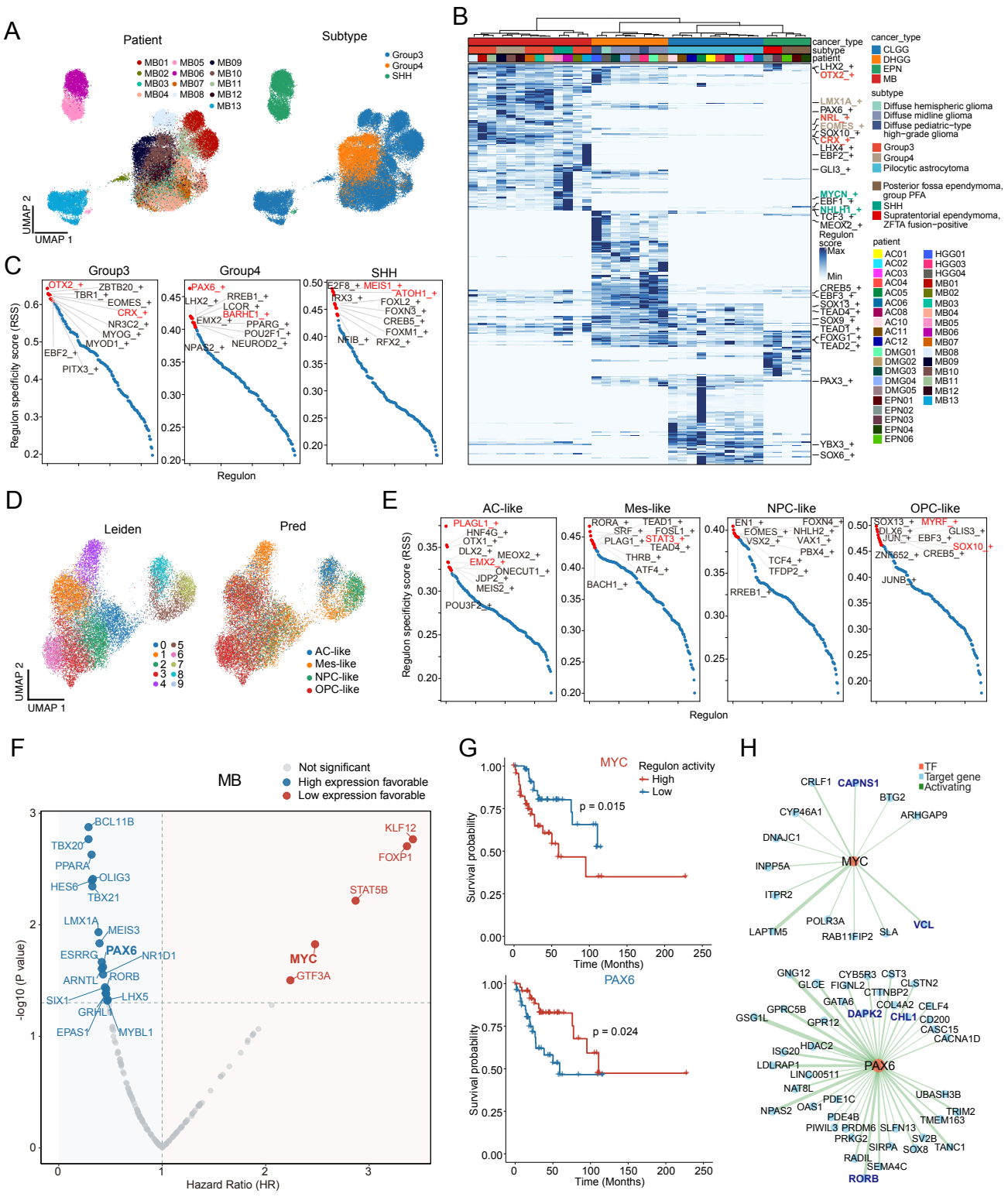

A
Patient
Subtype
UMAP 1
UMAP 2
B
cancer_type
subtype
patient
Regulon score
C
Group3
Group4
SHH
Regulon specificity score (RSS)
Regulon
D
Leiden
Pred
AC-like
Mes-like
NPC-like
OPC-like
E
AC-like
Mes-like
NPC-like
OPC-like
Regulon specificity score (RSS)
Regulon
F
MB
Not significant
High expression favorable
Low expression favorable
-log10 (P value)
Hazard Ratio (HR)
G
MYC
Regulon activity
High
Low
p = 0.015
PAX6
p = 0.024
Survival probability
Time (Months)
H
TF
Target gene
Activating
MYC
PAX6

Fig. 5

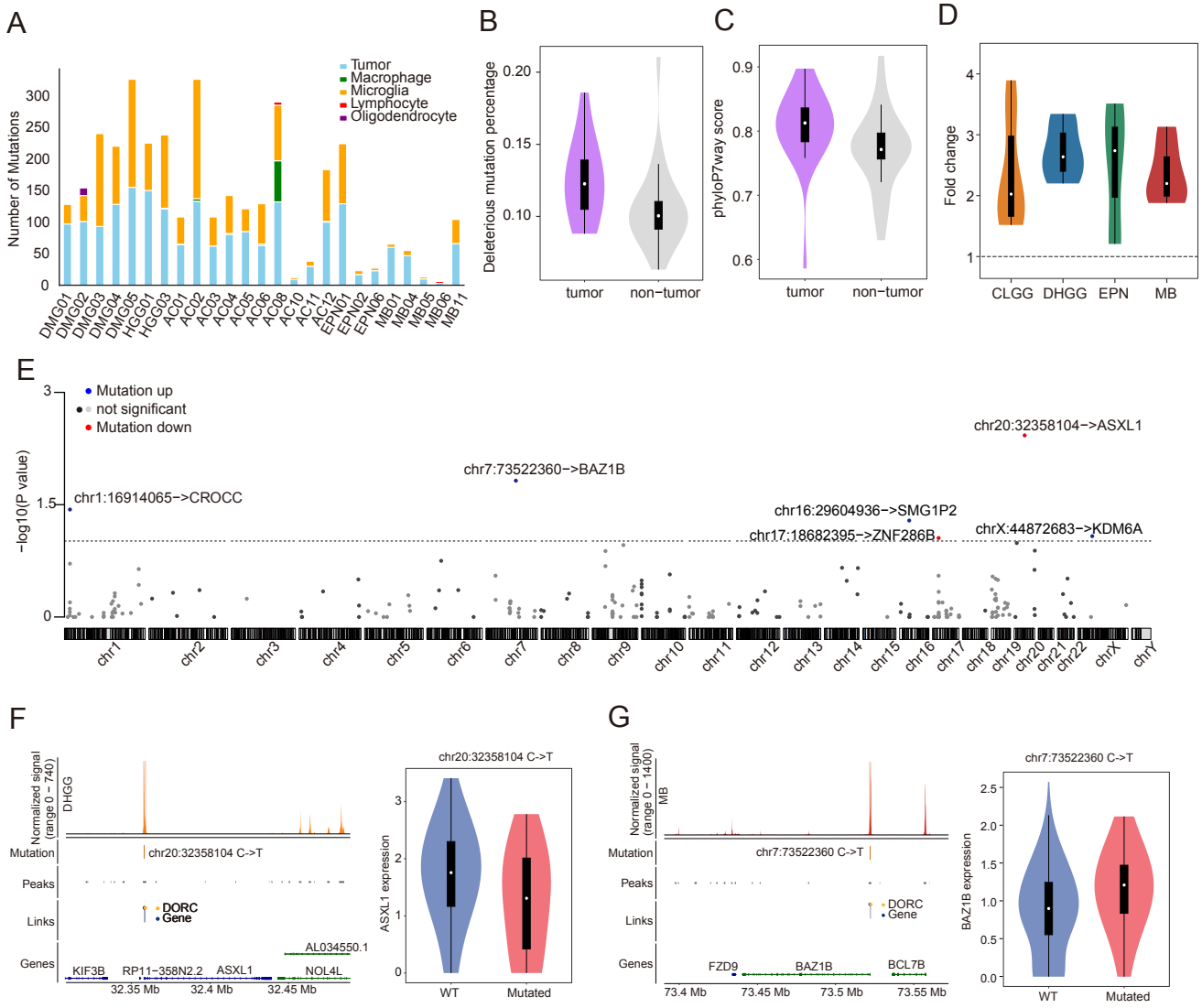
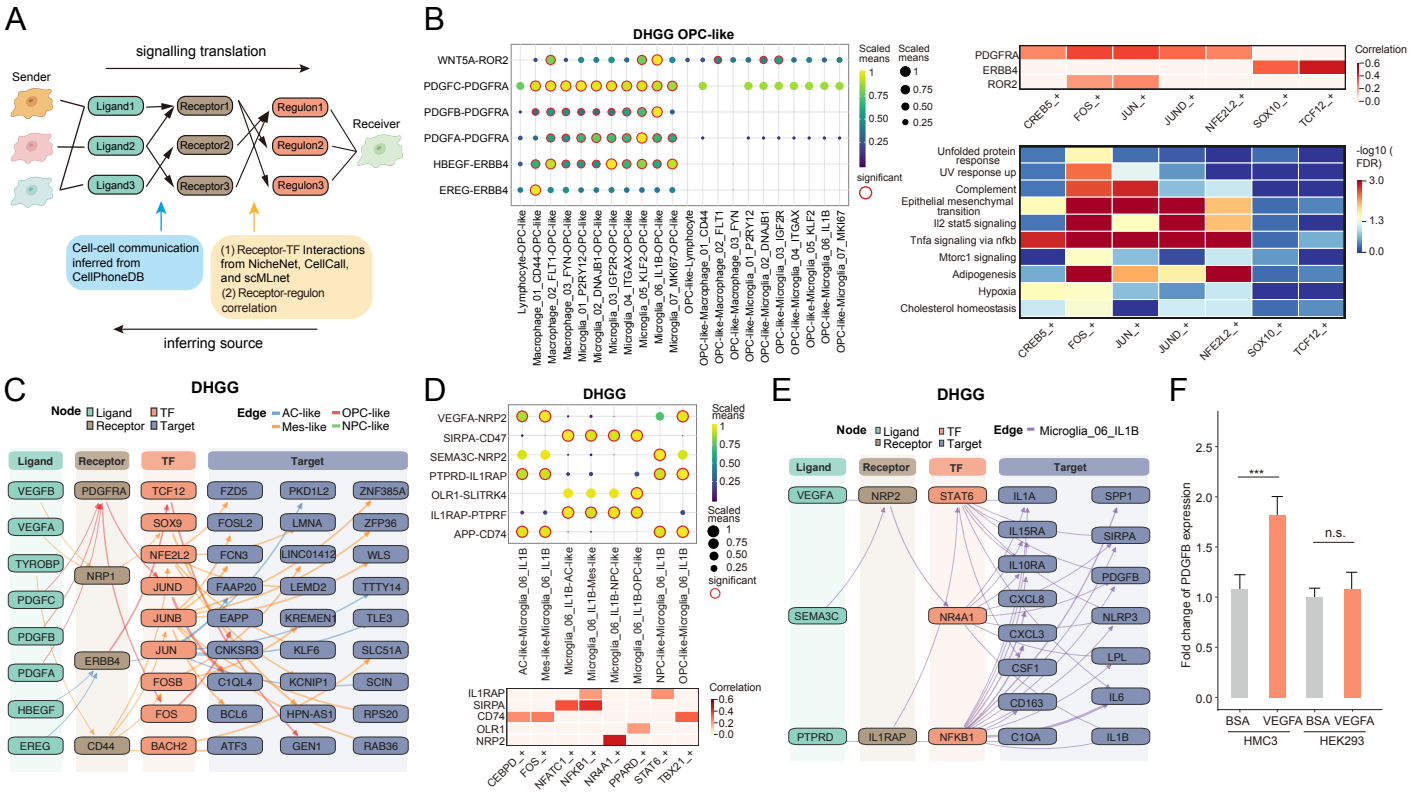

A
Number of Mutations
Tumor
Macrophage
Microglia
Lymphocyte
Oligodendrocyte
DMG01 DMG02 DMG03 DMG04 DMG05 HGG01 HGG03 AC01 AC02 AC03 AC04 AC05 AC06 AC08 AC10 AC11 AC12 EPN01 EPN02 EPN06 MB01 MB04 MB05 MB06 MB11
B
Deleterious mutation percentage
tumor
non−tumor
C
phyloP7way score
tumor
non−tumor
D
Fold change
CLGG DHGG EPN MB
E
Mutation up
not significant
Mutation down
−log10(P value)
chr20:32358104−>ASXL1
chr7:73522360−>BAZ1B
chr1:16914065−>CROCC
chr16:29604936−>SMG1P2
chr17:18682395−>ZNF286B
chrX:44872683−>KDM6A
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr13 chr14 chr15 chr16 chr17 chr18 chr19 chr20 chr21 chr22 chrX chrY
F
Normalized signal (range 0 − 740)
DHGG
Mutation
chr20:32358104 C->T
Peaks
Links
DORC Gene
Genes
KIF3B RP11−358N2.2 ASXL1 NOL4L AL034550.1
32.35 Mb 32.4 Mb 32.45 Mb
chr20:32358104 C->T
ASXL1 expression
WT
Mutated
G
Normalized signal (range 0 − 1400)
MB
Mutation
chr7:73522360 C->T
Peaks
Links
DORC Gene
Genes
FZD9 BAZ1B BCL7B
73.4 Mb 73.45 Mb 73.5 Mb 73.55 Mb
chr7:73522360 C->T
BAZ1B expression
WT
Mutated

Fig. 6


A
signalling translation
Sender
Ligand1
Ligand2
Ligand3
Receptor1
Receptor2
Receptor3
Regulon1
Regulon2
Regulon3
Receiver
Cell-cell communication inferred from CellPhoneDB
(1) Receptor-TF Interactions from NicheNet, CellCall, and scMLnet
(2) Receptor-regulon correlation
inferring source
B
DHGG OPC-like
WNT5A-ROR2
PDGFC-PDGFRA
PDGFB-PDGFRA
PDGFA-PDGFRA
HBEGF-ERBB4
EREG-ERBB4
Scaled means
significant
Correlation
PDGFRA
ERBB4
ROR2
CREB5_+
FOS_+
JUN_+
JUND_+
NFE2L2_+
SOX10_+
TCF12_+
Unfolded protein response
UV response up
Complement
Epithelial mesenchymal transition
Il2 stat5 signaling
Tnfa signaling via nfkb
Mtorc1 signaling
Adipogenesis
Hypoxia
Cholesterol homeostasis
-log10 (FDR)
C
DHGG
Node Ligand Receptor TF Target
Edge AC-like Mes-like OPC-like NPC-like
Ligand
Receptor
TF
Target
VEGFB
VEGFA
TYROBP
PDGFC
PDGFB
PDGFA
HBEGF
EREG
PDGFRA
NRP1
ERBB4
CD44
TCF12
SOX9
NFE2L2
JUND
JUNB
JUN
FOSB
FOS
BACH2
FZD5
FOSL2
FCN3
FAAP20
EAPP
CNKSR3
C1QL4
BCL6
ATF3
PKD1L2
LMNA
LINC01413
LEMD2
KREMEN1
KLF6
KCNIP1
HPN-AS1
GEN1
ZNF385A
ZFP36
WLS
TTTY14
TLE3
SLC51A
SCIN
RPS20
RAB36
D
DHGG
VEGFA-NRP2
SIRPA-CD47
SEMA3C-NRP2
PTPRD-IL1RAP
OLR1-SLITRK4
IL1RAP-PTPRF
APP-CD74
AC-like-Microglia_06_IL1B
Mes-like-Microglia_06_IL1B
Microglia_06_IL1B-AC-like
Microglia_06_IL1B-Mes-like
Microglia_06_IL1B-NPC-like
Microglia_06_IL1B-OPC-like
NPC-like-Microglia_06_IL1B
OPC-like-Microglia_06_IL1B
IL1RAP
SIRPA
CD74
OLR1
NRP2
CEBPD_+
FOS_+
NFATC1_+
NFKB1_+
NR4A1_+
PPARD_+
STAT6_+
TBX21_+
E
DHGG
Node Ligand Receptor TF Target
Edge Microglia_06_IL1B
Ligand
Receptor
TF
Target
VEGFA
SEMA3C
PTPRD
NRP2
IL1RAP
STAT6
NR4A1
NFKB1
IL1A
IL15RA
IL10RA
CXCL8
CXCL3
CSF1
CD163
C1QA
SPP1
SIRPA
PDGFB
NLRP3
LPL
IL6
IL1B
F
Fold change of PDGFB expression
***
n.s.
BSA
VEGFA
BSA
VEGFA
HMC3
HEK293