**IMPORTANT NOTE:** The following is a thorough annotation of a subset of clusters we identified within our transcriptional heterogeneity analysis project. Currently we are considering publishing this series of documents independently of the main paper, as an online resource for users to reference when diving into our pan-cancer atlas. These will be then updated with time as new and better information emerges. Alternatively, a stripped-down version could be published as supplemental information.

**T000 Central Nervous System**

The CNS tumours in our dataset exhibit three major divides: glioma vs. non-gliomas, gliomas with wildtype or mutated *IDH1*, and, in this last group, samples with and without hemizygous codeletion of chromosome arms 1p and 19q (Fig. 5a, Fig. S21a, b). We believe that the ability of our clustering system to demonstrate these genomic differences from purely transcriptional data with high confidence is a testament to its effectiveness. Past these major divides, differences in histotypes and tumour phenotype, as well as the transcriptional signals generated from specific mutations, become increasingly important to differentiate sibling classes. Indeed, many of these deeper clusters may reflect underlying clinical, genomic, and epigenomic differences which we are currently unable to determine due to a lack of relevant annotations. Our results express the plethora of undiscovered molecular subtypes of gliomas and glioblastomas in particular and highlight the need for further studies of subtype-specific drivers and their therapeutic susceptibilities.

**Medulloblastomas**

At the first level, we see the separation of medulloblastomas, in **T027 MLBLA** (n = 29), from the rest of CNS tumours **T026 CNS A** (n = 894) (Fig. S21b) Letters in the naming will be used in this setting to distinguish mixed classes that maintain the same composition of their parent class, with the removal of specific subtypes singled out into their sibling classes, as in this case. Interestingly, we also note the presence of a single pineal parenchymal tumour in **T026**. The recursion allows us to observe the emergence of known subtypes from literature at deeper level of this branch1. This class further splits into **T058 MLBLA G3/G4** (n = 37) (Fig. S21b), a cluster of mixed G3 and G4 subtypes, with overexpression (edgeR quasi-likelihood negative binomial generalized log-linear model (glmQLFTest)) of *OTX2* (logFC = 3.48, FDR = 6.368e-06) and *FOXG1* (logFC = 8.44, FDR = 4.026e-06), while **T059 MLBLA WNT/SSH** (n = 5), shows overexpression of both WNT (single-sample GSEA from GSVA2, median normalized enrichment score (medNES) = 1.15, Mann Whitney *U* test Benjamin-Hochberg adjusted p-value (MWU adj. p-val) = 6.74e-05) and SHH (medNES = 1.42, MWU adj. p-val = 2.02e-04) pathways3. While samples of the G3 and G4 subtypes are then separated at the next level into **T060 MLBLA G4** (n = 9) and **T061 MLBLA G3** (n = 15), the population of **T063** is well below our set cut-off, preventing RACCOON from dividing WNT and SHH subtypes. **T060** overexpresses genes of the G4 subtype, including SNCAIP (logFC = 5.68, FDR = 1.11e-05), DIRAS3 (logFC = 4.35, FDR = 2.351e-06), KCNA1 (logFC = 4.19, FDR = 3.684e-04), and RND1 (logFC = 3.26, FDR = 1.542e-04), while **T61** overexpresses genes upregulated in the G3 subtype, such as PDE6H (logFC = -6, FDR = 6.038e-04), GNGT1 (logFC = -6.1, FDR = 2.651e-04), and NPR3 (logFC = -5.71, FDR = 4.824e-04).

**Separation by IDH1 status**

Following the remainder of CNS tumours after the removal of medulloblastomas, we observe the separation of gliomas without *IDH1* mutations, which form **T028 CNS IDHwt** (n = 406) from samples with *IDH1* mutations (19/222 vs 417/433, χ2 p-val < 2.2e-16), which form **T029 CNS IDHmut** (n = 488) (Fig. S21a, b). The latter has patients with lower median age (49.00 vs 38.00 y.o., MWU p-val = 2.04e-3), but **T028** has a considerably higher proportion of paediatric patients (40.06% vs. 27.05%, χ2 p-val = 2.40e-05). Furthermore, **T028** displays patients with significantly worse survival (Kaplan Meier log rank test (lrt) p-val = 1.57e-50 at 6423 days) in line with literature4 reaching median overall survival (OS) at only 448 days compared to **T029** at 2907 (Fig. S21c).5–78

**BCOR Altered Samples and Ependymomas**

Along the *IDH1* wild-type branch **T028** we then observe the separation of gliomas and glioblastomas in **T030 GLI IDHwt** from ependymomas in **T032 EPDY** and samples with lesions of the BCL-6 corepressor protein gene, *BCOR* **T031 CNS BCOR/PNET** (Fig. 5a, Fig. S21a, b). **T030 GLI IDHwt** (n = 364) includes the vast majority of gliomas and glioblastomas without mutations of *IDH1* and is the oldest class (median age = 52 y.o. KW p-val = 5.72e-10).

**T031 CNS BCORmut/PNET** is a peculiarly small cluster (n = 12) comprised of heterogeneous diagnoses. It includes a variety of brain and CNS tumours, including ependymomas, primitive neuroectodermal tumours (PNET), gliomas, an embryonal tumour with multi-layered rosettes, and a handful of solid tumours - several possibly misdiagnosed as Ewing sarcoma - and one infantile fibrosarcoma (Fig. 5d). All samples are from paediatric patients, with a median age of 4.5 y.o. This cluster is characterized by an overexpression of *BCOR* (median logFC= 4.38, FDR ≤ 2.94e-41) (Fig. 5c).

*BCOR* participates in a range of chromatin altering activities including binding to histone acetylases and chromatin-altering complexes, namely polycomb group complexes9. Alterations of these genes, many of which consist of fusions or internal tandem duplications (ITD) (Fig. 5b), have been well characterized in both soft tissue tumours and a recently defined group of CNS neoplasms: high grade neuroepithelial tumours of the central nervous system (CNS HGNET-BCOR)10. Gene set enrichment analyses revealed significant upregulation of both WNT (medNES ≥ 1.35, Kruskal–Wallis one-way analysis of variance test Benjamin-Hochberg adjusted p-value (KW adj. p-val) = 3.83e-09, Dunn’s Test of Multiple Comparisons Benjamin-Hochberg adjusted p-value (Dunn adj. p-val) < 1.00e-04)11 and SHH (medNES ≥ 1.51, KW adj. p-val= 6.16e-09, Dunn adj. p-val < 1.00e-04) pathways12, as well as basal cell carcinoma pathways (medNES ≥ 1.70, KW adj. p-val = 2.84e-20, Dunn adj. p-val < 1.00e-04)3 in line with what is reported in literature. We also observe significant overexpression of *NTRK3* (median logFC= 2.45, FDR ≤ 1.7e-16), but not *NTRK2* (FDR ≤ 5.847e-01) and *NTRK1* (FDR ≤ 9.063e-01) in **T031** vs. **T030** and **T032**, as commonly described in *BCOR*-ITD sarcomas13,14.

Finally, **T032 EPDY** (n = 30) is comprised almost exclusively of ependymomas. It is the cluster with the youngest patients, with a median age of 2.64 y.o. No subtypes are identified, possibly due to the limits in the reference dataset population.

**IDH wild-type gliomas**

At the next level, we observe the separation between a small paediatric cluster **T033 GLI LG PED** (n = 63) and a much larger adult class **T034 GLI HG** (n=301) (Fig. S21b, d). Both contain mixed diagnoses but with a strong majority of samples labelled as gliomas. There’s a significant difference in age, with **T033** having a population with a median age of only 9.00 y.o. versus **T0**34 with 56.00 y.o. (MWU p-val 4.00e-20). **T034** is characterized by significant upregulation of *HOX* genes (36/39 FDR < 0.05), particularly *HOXD9* (logFC = -5.03, FDR = 1.20e-23) and *HOXA5* (logFC = -6.18, FDR = 2.40e-29)7,15 which have been associated with cancer cell survival and proliferation in gliomas. Together with overexpression of *VEGFA*16 (logFC = -1.04, FDR = 2.15e-05), a marker of poor survival, and glioma stemness genes *TERT* and *EGFR* (FDR ≤ 1.00e-28) 17, this profile suggests **T034** to be a class of high-grade gliomas and glioblastoma multiforme, while **T033** to be a largely paediatric, low-grade glioma class – though all samples from the TCGA are astrocytomas (6/6 vs 50/232, χ2 p-val = 6.74e-05). This is supported by **T033** being enriched for grade II (3/6 vs 10/232, χ2 p-val = 7.73e-05) samples, with **T034** being enriched for grade IV samples (0/6 vs 155/232, χ2 p-val = 3.11e-03). However, we are unable to confirm differences in survival due to a lack of clinical annotation of samples in **T033.**

The glioma subtypes run much deeper along complex hierarchical paths. At the next level, **T034** splits into **T035 GLI HG LOH c7/10** (n =236) and **T036 GLI HG PRON** (N = 65) (Fig. S21d). Both are mixed glioma and glioblastoma groups. We also observe a significant difference in age (median 58.00 vs 35.00 y.o. MWU p-val = 8.76e-06) and paediatric composition (13.56% vs 50.77%, χ2 p-val = 3.27e-10). There is no difference in overall survival between the groups (lrt p-val = 8.23e-02 at 6423 days)18. **T035** contains almost all samples of the classical (85/185 vs. 1/29, χ2 p-val = 3.527e-05) and mesenchymal (87/185 vs. 3/29, χ2 p-val = 4.343e-04) expression subtypes, while **T036** is almost wholly composed of the proneural subtypes (2/185 vs. 24/29, χ2 p-val < 2.2e-16); although the majority of neural type samples are also found in **T035**, the difference is not significant (11/185 vs. 1/29, χ2 p-val = 0.9128)18,19. **T035** shows significant overexpression of *SAA1* (logFC = 4.84, FDR = 2.869e-16), *MEOX2* (logFC = 4.79, FDR = 8.46e-22), *CHI3L1* (logFC = 3.5, FDR = 6.93e-20)*, S100A4* (logFC = 2.04, FDR = 1.26e-18)and *ANXA1* (logFC = 2.68, FDR = 1.18e-37), all associated with poor survival16,20 ,and has a considerably higher leukocyte content than **T036** (0.190 vs. 0.059, MWU p-val = 1.42e-08)21. In turn, **T036** samples overexpress *PDGFRA* (logFC= -2.8, FDR = 3.80e-34), a marker of the proneuronal expression type19. **T035** contains more *TP53* mutants (χ2 p-val =2.11-02), and is also enriched for genesets concerning loss of heterozygosity (LOH) of regions implicated in gliomagenesis (medNES = 1.32, MWU adj. p-val = 2.15e-06)22, suggesting it contains samples with gain of chromosome 7 and loss of chromosome 10. This is further supported by its overexpression of *EGFR* (logFC = 3.47, FDR = 1.18e-18) and is in line with literature, in which classical GBM samples tend to harbour these lesions. Indeed, **T035** is highly enriched for tumours with gain chr7/loss chr10, confirmed by clinical data (139/200 vs. 15/35, χ2 p-val = 4.146e-03)18. **T036** contains a greater proportion of *ATRX* mutant tumours (9/194 vs. 10/24, χ2 p-val = 7.31e-06)18.

Glioblastomas and high-grade gliomas separate at the next level within **T036** (Fig. S21b)**.** We observe **T042 GLI HG/GBM PRON** (n = 48) carrying glioblastomas mostly of the proneuronal subtype and **T043 GLI HG PED H3.3mut** (n =19) with the rest of the samples, primarily marked as high-grade gliomas from St. Jude’s (χ2 p-val = 8.75e-14) (Fig. S21e). The two classes also differ significantly in age, with **T042** having patients with a median age of 44.5 y.o. while **T043** has a median age of 5.85 y.o. (MWU p-val 3.88e-05). In fact, **T043** is the cluster with the youngest group of patients within the entire cohort of both gliomas and gliobastomas and is one of only two clusters with >90% paediatric composition, the other being **T033** GLI LG. Given that it’s a majority paediatric cluster whose parent cluster demonstrates very poor survival, **T043** may represent H3.3 (*H3F3A*) mutated tumours. Support for this hypothesis comes from enrichment of gene sets involving H3.3 mutation (here nominally *K27M*) between **T043** and **T042** (medNES = 1.73, MWU adj. p-val = 3.39e-02)23. Notch signalling (medNES = 1.05, MWU adj. p-val = 2.31e-02) and neural differentiation (medNES = 1.04, MWU adj. p-val =6.10e-03)24 genesets are also enriched in T043 and are a feature of these tumours. 25,26 Going back up along the hierarchy, **T035 GLI HG LOH c7/10** also splits in two (Fig. S21b), with **T037 GLI HG NEUR** **DIFFhigh** (n = 93) being comprised of gliomas and glioblastomas of the classical (49/68 vs. 36/117, χ2 p-val = 1.29e-07) and neural subtypes (8/68 vs. 3/117, χ2 p-val = 2.58e-02) and **T038 GLI HG/GBM MES/CLASS** (n = 143) carrying a mixture of glioblastomas multiforme subtypes. **T037** is composed of a majority of astrocytomas (34/78 vs. 10/118), χ2 p-val = 3.13e-08) while **T038** contains a majority of glioblastomas (31/78 vs. 100/118, χ2 p-val = 9.30e-11). Interestingly, while almost all of the gliomas in **T038** are marked as *IDH1* wild type, a handful of samples are *IDH1* mutant (0/84 vs. 10/114, χ2 p-val = 1.40e-2), suggesting these may be passenger rather than driver mutations. **T038** also has significantly higher leukocyte fraction (0.178 vs. 0.248, MWU, p-val = 1.77e-02). There are no differences in proportion of gain chr7/loss chr10 samples (χ2 p-val = 3.80e-01).

The remaining subtypes are found in the child classes of **T038** (Fig. S21b)**:** we find the majority of classical samples (31/34 vs 2/27, vs 1/43, χ2 p-val < 2.2e-16) in **T039 GLI HG/GBM CLASS** (n = 37), mesenchymal subtype samples in both **T040 GLI HG/GBM MES** (n = 36) and **T041 GLI HG/GBM NEUR ATRXmut** (n=57) (3/34 vs. 24/27 vs. 38/43, χ2 p-val = 3.28e-14). **T040** contains two concurrent *PIK3CA* and *NF1* mutated samples (χ2 p-val = 2.73e-02), *NF1* mutations are typical of mesenchymal GBMs27. **T041** inherits all *IDH1* mutants (0/34 vs. 0/26 vs. 10/42, χ2 p-val = 3.64e-04) and is enriched for *TP53* mutants (0/5 vs. 0/5 vs 5/6, χ2 p-val =1.38e-2). Patients at **T039** have the best overall survival, reaching median OS at 375 days post diagnosis, while those in **T040** have the worst, reaching median OS at 225 days (lrt p-val = 3.44e-02 at 2549 days). These clusters differ in their share of *TERT* promoter mutations and *ATRX* mutations when available, respectively, with **T039** and **T040** comprised of samples with *TERT* promoter mutants (8/8 vs. 9/9 vs. 4/9, χ2 p-val = 2.89e-03) while **T041** contains all *ATRX* mutants (0/33 vs. 0/25 vs. 7/40, χ2 p-val = 4.23e-03). Examination of telomere maintenance pathways reveals samples with relevant data in **T041** to be driven more by *ATRX* mutations (0/8 vs. 0/8, 4/5, χ2 p-val =1.4513-02) while its siblings are wholly driven by *TERT* mutations (8/8, 8/8, 4/5, χ2 p-val = 3.87e-03). Analysis of gene sets for relevant pathways shows **T040** to be enriched for mesenchymal GBM over its siblings (medNES ≥ 1.27, KW adj. p-val = 8.91e-14, Dunn adj. p-val < 1e-04), while **T041** is enriched for neural GBM (medNES ≥ 1.38, KW adj. p-val = 5.59e-14, Dunn adj. p-val < 1e-04), suggesting this subtype has a more neural than mesenchymal identity This is further supported by **T041’**s inheritance of the majority of neural (0/34 vs. 1/27 vs. 2/43, χ2 p-val = 4.60e-01) and proneural samples (0/34 vs. 0/27 vs. 2/43, χ2 p-val = 2.35e-01), though neither reach significance.

While it is surprising to see two unrelated clusters of glioblastomas containing large populations of classical expression subtype glioblastomas, **T037** and **T039,** closer examination reveals **T039** to represent a bona fide classical GBM subtype, supported by significant enrichment of classical glioblastoma gene signatures (medNES 1.15, MWU adj. p-val = 1.61e-08), and by its higher proportion of classical samples (31/34 vs. 49/68, FET p-val = 3.94e-02) (Fig. S21e). **T037** is instead enriched for neural (medNES = 1.79, MWU adj. p-val = 1.87e-13) and proneural signatures (medNES = 1.15, MWU adj. p-val = 8.81e-05) over **T039** (Fig. S21e)**.** Furthermore, the presence of a sizeable astrocytoma/glioma component in **T037** suggests it may represent a more “mixed” phenotype of GBM/glioma than its sibling **T038,** whose children separate into histotype-specific component clusters of GBM. Indeed, **T037** is enriched for neural and proneural signatures against all children of **T038** (medNES ≤ 1.04, KW adj. p-val ≤ 4.07e-16) (Fig. S21e). We hypothesize **T037** represents a more neurally differentiated class, transcending canonical subtyping. This is further supported by enrichment of genes pertaining to neural development (medNES = 1.06, KW adj. p-val = 7.01e-14, Dunn adj. p-val < 0.05) and differentiation (medNES = 1.11, KW adj. p-val = 2.24e-12, Dunn post-hoc adj. p-val < 0.001)28,29 in **T037** with respect to **T039,** **T040,** and **T041** (Fig. S21e)**.**

**IDH mutant gliomas**

Along the alternative branch hosting IDH mutant gliomas (**T029**) we find that the hemizygous codeletion of chromosome arms 1p and 19q is a major driver in the separation of classes: we find samples with codeletion in **T044 GLI IHDmut CODEL** (n = 270) and samples without codeletion in **T045 GLI IDHmut noCODEL** (n = 218) (168/ 221 vs. 1/213, χ2 p-val < 2.2e-16) (Fig. S21b, c). Though **T044** has significantly older patients (median age 40 vs. 36 y.o. MWU p-val = 2.56e-03), it has a slightly larger paediatric population (28.14% vs. 25.69%). According to the clinical information from TGCA, **T044** contains significantly more tumours with *TERT* promoter mutations (89/125 vs. 6/124, χ2 p-val < 2.2e-16), while **T045** is enriched for *ATRX* mutants (23/ 221 vs. 155/211, χ2 p-val < 2.2e-16). **T044** contains a majority of oligodendrogliomas (134/198 vs. 28/183, χ2 p-val < 2.2e-16) while **T045** contains a majority of astrocytomas (17/198 vs. 97/183, χ2 p-val < 2.2e-16). However, despite differences in codeletion status, we find no difference in overall survival between the two groups (lrt p-val = 3.78e-01 at 5546 days).

Following along **T044,** we observe the singling out of a small set of low-grade gliomas (n = 12/30 vs 4/240, χ2 p-val = 1.54e-15) and dysembryoplastic neuroepithelial tumours (DNET) (10/30 vs. 2/240, χ2 p-val = 1.67e-14) in **T046 GLI LG IDHmut CODEL/DNET** (n = 30) from the rest of the gliomas in **T047 GLI IHDmut CODEL A** (n = 240)**.** There is a significant age disparity between the two clusters (14.11 vs. 41 y.o., MWU p-val = 9.61e-11) as the former class is made up entirely of paediatric samples.

**T047** further split by age and histotype. **T048 GLI IHDmut** **MULTICELL** **NET** (n = 67) has significantly younger patients than **T049 GLI IHDmut CODEL B** (n = 173) (median 35.00 vs 44.00 y.o. MWU p-val = 2.26e-03) due to its larger paediatric component (38.80% vs. 11.56%, χ2 p-val = 3.70e-06). There is no difference in overall survial (lrt p-val = 6.23e-02 at 5546 days). While **T048** contains more astrocytomas (13/48 vs 4/150, χ2 p-val = 7.06e-07), **T049** has a considerably higher oligodendroglioma population (20/48 vs 114/150, χ2 p-val = 2.14e-05). According to TCGA data, **T048** is enriched for *EGFR* (3/33 vs. 0/87, χ2 p-val = 2.83e-02), *ATRX* (12/33 vs. 6/87, χ2 p-val = 1.77e-04) and *TP53* (13/33 vs. 7/87, χ2 p-val =1.23e-04) mutant tumours, while **T049** contains more *CIC* (χ2 p-val = 1/33 vs. 48/87, 6.33e-07), *FUBP1* (1/33 vs. 22/87, χ2 p-val = 1.22e-02) and *NOTCH1* (0/33 vs. 19/87, χ2 p-val =8.14e-03) mutants. Most tumours in **T048** are neural (46/49 vs. 38/144, χ2 p-val = 7.39e-16), while the majority of those in **T049** are proneural (2/49 vs. 105/144, χ2 p-val = 2.26e-16) Most importantly, and quite unexpectedly, **T048** is mostly composed of *IDH1*wt (16/53) and non-codeleted samples (45/53). It is not clear why this class is found within the *IDH1*codeleted branch.

Gene set enrichment analysis reveals that every locus available for chr1p - with the expectation of chr1p11 - (MWU p-val ≤ 8.35e-04) and chr19q (MWU p-val ≤ 5.18e-23) are significantly downregulated in **T049** compared to **T048,** confirming more severe population-wide loss of these loci in **T049** vs. **T048** and supporting that, true to their annotation, the majority of samples in **T048** have normal expression of these loci, in spite of their transcriptional similarities with the codeleted branch. The overall expression profile of both IDH wild type and non-codeleted tumours within **T048** have a high correlation with true chr1p/19q co-deleted IDHmut gliomas within **T044** (R ≥ 0.802, Pearson correlation p-val < 2.20e-16).

Further examination of gene sets upregulated in **T048** compared to its sibling class **T049,** its uncle class **T045** GLI IDHmut noCODEL, and its cousin class **T030 GLI IDHwt** revealed significant upregulation of genesets related to neuron-neuron synaptic transmission (KW adj. p-val = 2.35e-89, medNES = 1.13, Dunn adj. p-val-val < 1.00e-04), synaptic plasticity (KW adj. p-val = 5.23e-85, medNES = 1.30, Dunn adj. p-val-val < 1.00e-04), neurite formation (KW adj. p-val = 2.45e-51, medNES = 1.15, Dunn adj. p-val < 1.00e-04)30, and microtubule polymerization (KW adj. p-val = 1.59e-63, medNES = 1.10, Dunn adj. p-val-val < 1.00e-04) (Fig. S21e). We also observe upregulation of glutaminergic signalling (KW adj. p-val = 1.06e-102, medNES = 1.31, Dunn adj. p-val< 1.00e-04), particularly of AMPA cationic channel activity (KW adj. p-val = 9.08e-59, medNES = 1.43, Dunn adj. p-val < 1.00e-04) – including AMPA-dependent synaptic plasticity (KW adj. p-val = 7.36-84, Dunn adj. p-val < 1.00e-04), and of extracellular calcium export (KW adj. p-val = 7.61e-93, Dunn adj. p-val < 1.00e-04, medNES = 1.14)30–32 (Fig. S21e).

We also observe increases in gap junction formation (KW adj. p-val = 5.81e-35, medNES = 2.39, Dunn adj. p-val < 1.00e-04) and connexin binding (KW adj. p-val = 3.61e-28, medNES = 1.26, Dunn adj. p-val < 1.00e-04) (Fig. S21e).

Taken together, these results suggest **T048** to be composed of gliomas of a recently described multicellular network phenotype, a pro-invasive and radioresistant resistant mode of glioma growth33. Gene expression analysis reveals significant upregulation of *GAP34* in **T048** vs. other *IDH1* mutant tumour groups (**T045** and **T049**) (median logFC = 1.33, FDR ≤ 1.80e-13), the principal gap-junction protein mediating this phenotype33, as well as *NOTCH1* underexpression (median logFC= -1.16, FDR ≤ 1.950e-06) and downregulation of *NOTCH1* signalling (KW adj. p-val = 1.65e-45, medNES = 0.92, Dunn adj. p-val-val < 0.05)30 over all other glioma types, the crucial determinant of this phenotype34 (Fig. S21e). This is despite the lack of *NOTCH1* mutant samples in **T048**; T049 contains the majority *NOTCH1* mutants of the glioma cohort (vs. **T048**, **T045**,and **T030**, 0/33 vs. 19/87 vs. 4/113 vs. 0/52, χ2 p-val = 2.29e-07) and exhibits the highest *NOTCH1* expression (median logFC= 0.83, FDR ≤ 2.961e-02), so we speculate these *NOTCH1* mutations to be gain-of-function. However, despite this phenotype displaying radioresistance, samples in **T048** show no significant differences in overall survival compared to other IDH mutated glioma groups (**T045** and **T049**) at 6423 days.

We speculate this novel phenotype may have good transcriptional affinity with chr1p/19q codeletion, in spite of the lack of apparent lesions in the region.

Though this phenotype is mostly associated with astrocytomas33, **T048** is a mixed cluster – containing large amounts of both astrocytomas and oligodendrogliomas. **T048** then splits in two classes (Fig. S21b), with different histological populations; **T050 GLI IDHmut MULTICELL NET OLIGOD** (n = 31) contains more oligodendrogliomas (15/22 vs. 5/26, χ2 p-val = 1.73-3) than **T051 GLI IDHmut MULTICELL NET ASTROC** (n = 36), which instead is populated by astrocytomas (0/22 vs. 13/26, χ2 p-val = 3.74e-04).3536**T050** also inherits the majority of chr1p/19q codelted samples (7/16 vs. 1/29, χ2 p-val = 1.91e-02).

Similarly, **T049** splits by histological composition (Fig. S21b) with **T052 GLI IHDmut CODEL NOTCH1** (n = 89) being enriched (68/81 vs. 46/69, χ2 p-val = 2.27e-02) for oligodendrogliomas and **T053 GLI IHDmut CODEL OLIGOAST** (n = 84) for oligoastrocytomas (11/81 vs. 21/69, χ2 p-val = 2.08e-02). **T052** also has a significantly higher population of *NOTCH1* mutant samples (17/56 vs. 2/39, χ2 p-val = 2.07e-02). **T051** contains a larger share of neural gliomas (30/78 vs. 8/66, χ2 p-val = 7.154e-04), while **T052** contains more proneural gliomas (48/78 vs. 57/66, χ2 p-val =1.623-03).

Finally, following along the non-codeleted branch defined by **T045,** we observe four children classes, characterized by significant differences in the sex ratios (Fig. S21b): **T054 GLI** **IDHmut noCODEL OLIGOC** (n = 99) is composed by 63.64% of males, **T055 GLI IDHmut noCODEL NEUR DIFFhigh** (n = 30) is entirely female (χ2 p-val = 2.65e-17), **T056 GLI IDHmut noCODEL MES** (n = 24) is 75.00% male and **T057 GLI IDHmut noCODEL H3demet** (n = 47) is almost exclusively male (97.87% χ2 p-val = 2.65e-17). Although sex differences have previously been reported to be associated with differences in survival in glioma and GBM37, patients in these clusters have no significant differences in overall survival (lrt p-val = 1.44e-01 at 4752 days).

Differential expression and gene sets analyses on these classes revelated that **T054** is enriched for oligodendrocyte development (medNES = 1.03, adj p-val = 4.85e-15) and myelination (med NES = 1.23, KW adj. p-val = 3.97e-09, Dunn post-hoc test p < 0.01)28,29 and overexpresses *MBP* and *MOBP* (FDR ≤ 2.461e-04), predictors of improved survival38,39. **T055** is enriched for gene sets related to neuronal development (medNES ≥ 1.01, KW adj. p-val = 9.36e-17). **T056** is enriched for genesets involving *MYC* signalling (medNES ≥ 1.04, KW adj. p-val ≤ 6.76e-03, Dunn adj. p-val < 0.05), the G2M checkpoint,(medNES ≥ 1.16, KW adj. p-val = 1.46e-03, Dunn adj. p-val < 0.01), and the immune response (medNES ≥ 1.510, KW adj. p-val ≤ 1.05e-05, Dunn adj. p-val < 0.01)40; it also exhibits the highest immune infiltration score out of its siblings (median = 990 vs. 648 vs. 2185 vs. 977, Dunn adj. p-val ≤ 4.81e-05). It overexpresses *MMP9* (median logFC = 1.41, FDR ≤ 3.515e-02), *CHI3L1* (median logFC = 1.79, FDR 8.332e-03), *S100A4* (median logFC= 2.12, FDR ≤ 4.281e-09), *EN1* (median logFC= 5.01, FDR ≤ 8.643e-14), and *ANXA1* (median logFC= 2.56, FDR ≤ 7.266e-14), markers of poor prognosis, and *IGF2BP3* (median logFC= 4.06, FDR ≤ 8.147e-13), a GBM-specific proliferative and invasive marker. **T056** is also the only cluster to contain a significant population of mesenchymal samples (0/58 vs. 0/21 vs. 5/11 vs. 2/ 35, χ2 p-val = 3.211e-08) and is enriched for epithelial mesenchymal transition genesets (medNES ≥ 1.15, KW adj. p-val = 4.46e-07, Dunn adj. p-val < 0.05)40. **T057** is enriched for genesets involving H3K4 demethylation (medNES ≥ 1.07, KW adj. p-val ≤ 6.91e-11, Dunn adj. p-val < 0.05) and H3K27 demethylation (medNES ≥ 1.08, KW adj. p-val ≤ 6.00e-03)28,29. It also overexpresses *LDHC* (median LogFC= 3.44, FDR ≤ 7.16e-06), which was found to be elevated in mesenchymal glioma stem cells and negatively correlates with survival41,42.

**Bibliography**

1. Northcott, P. A., Dubuc, A. M., Pfister, S. & Taylor, M. D. Molecular subgroups of medulloblastoma. *Expert Rev Neurother* **12,** 871–884 (2012).

2. Hänzelmann, S., Castelo, R. & Guinney, J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* **14,** 7 (2013).

3. Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28,** 27–30 (2000).

4. Hartmann, C. *et al.* Patients with IDH1 wild type anaplastic astrocytomas exhibit worse prognosis than IDH1-mutated glioblastomas, and IDH1 mutation status accounts for the unfavorable prognostic effect of higher age: implications for classification of gliomas. *Acta Neuropathol.* **120,** 707–718 (2010).

5. Park, S. Y., Piao, Y., Jeong, K. J., Dong, J. & de Groot, J. F. Periostin (POSTN) regulates tumor resistance to antiangiogenic therapy in glioma models. *Mol. Cancer Ther.* **15,** 2187–2197 (2016).

6. Steponaitis, G. *et al.* High CHI3L1 expression is associated with glioma patient survival. *Diagn Pathol* **11,** 42 (2016).

7. Cimino, P. J. *et al.* Increased HOXA5 expression provides a selective advantage for gain of whole chromosome 7 in IDH wild-type glioblastoma. *Genes Dev.* **32,** 512–523 (2018).

8. Hernández, M., Martín, R., García-Cubillas, M. D., Maeso-Hernández, P. & Nieto, M. L. Secreted PLA2 induces proliferation in astrocytoma through the EGF receptor: another inflammation-cancer link. *Neuro. Oncol.* **12,** 1014–1023 (2010).

9. Gearhart, M. D., Corcoran, C. M., Wamstad, J. A. & Bardwell, V. J. Polycomb group and SCF ubiquitin ligases are found in a novel BCOR complex that is recruited to BCL6 targets. *Mol. Cell. Biol.* **26,** 6880–6889 (2006).

10. Sturm, D. *et al.* New Brain Tumor Entities Emerge from Molecular Classification of CNS-PNETs. *Cell* **164,** 1060–1072 (2016).

11. Liberzon, A. *et al.* Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27,** 1739–1740 (2011).

12. Nishimura, D. BioCarta. *Biotech Software & Internet Report* **2,** 117–120 (2001).

13. Kao, Y.-C. *et al.* BCOR-CCNB3 Fusion Positive Sarcomas: A Clinicopathologic and Molecular Analysis of 36 Cases With Comparison to Morphologic Spectrum and Clinical Behavior of Other Round Cell Sarcomas. *Am. J. Surg. Pathol.* **42,** 604–615 (2018).

14. Kao, Y.-C. *et al.* NTRK3 overexpression in undifferentiated sarcomas with YWHAE and BCOR genetic alterations. *Mod. Pathol.* **33,** 1341–1349 (2020).

15. Tabuse, M. *et al.* Functional analysis of HOXD9 in human gliomas and glioma cancer stem cells. *Mol. Cancer* **10,** 60 (2011).

16. Xu, Y. *et al.* Screening critical genes associated with malignant glioma using bioinformatics analysis. *Mol. Med. Rep.* **16,** 6580–6589 (2017).

17. Beck, S. *et al.* Telomerase activity-independent function of TERT allows glioma cells to attain cancer stem cell characteristics by inducing EGFR expression. *Mol. Cells* **31,** 9–15 (2011).

18. Ceccarelli, M. *et al.* Molecular profiling reveals biologically discrete subsets and pathways of progression in diffuse glioma. *Cell* **164,** 550–563 (2016).

19. Brennan, C. W. *et al.* The somatic genomic landscape of glioblastoma. *Cell* **155,** 462–477 (2013).

20. Tachon, G. *et al.* Prognostic significance of MEOX2 in gliomas. *Mod. Pathol.* **32,** 774–786 (2019).

21. Thorsson, V. *et al.* The immune landscape of cancer. *Immunity* **48,** 812–830.e14 (2018).

22. Roversi, G. *et al.* Identification of novel genomic markers related to progression to glioblastoma through genomic profiling of 25 primary glioma cell lines. *Oncogene* **25,** 1571–1583 (2006).

23. Larson, J. D. *et al.* Histone H3.3 K27M accelerates spontaneous brainstem glioma and drives restricted changes in bivalent gene expression. *Cancer Cell* **35,** 140–155.e7 (2019).

24. Schaefer, C. F. *et al.* PID: the pathway interaction database. *Nucleic Acids Res.* **37,** D674-9 (2009).

25. Paugh, B. S. *et al.* Genome-wide analyses identify recurrent amplifications of receptor tyrosine kinases and cell-cycle regulatory genes in diffuse intrinsic pontine glioma. *J. Clin. Oncol.* **29,** 3999–4006 (2011).

26. Chen, K.-Y. *et al.* Reciprocal H3.3 gene editing identifies K27M and G34R mechanisms in pediatric glioma including NOTCH signaling. *Commun. Biol.* **3,** 363 (2020).

27. Fadhlullah, S. F. B. *et al.* Pathogenic mutations in neurofibromin identifies a leucine-rich domain regulating glioma cell invasiveness. *Oncogene* **38,** 5367–5380 (2019).

28. The Gene Ontology Consortium. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res.* **47,** D330–D338 (2019).

29. Ashburner, M. *et al.* Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25,** 25–29 (2000).

30. Jassal, B. *et al.* The Reactome Pathway Knowledgebase. *Nucleic Acids Res.* **48,** D498–D503 (2020).

31. Venkataramani, V. *et al.* Glutamatergic synaptic input to glioma cells drives brain tumour progression. *Nature* **573,** 532–538 (2019).

32. Venkatesh, H. S. *et al.* Electrical and synaptic integration of glioma into neural circuits. *Nature* **573,** 539–545 (2019).

33. Osswald, M. *et al.* Brain tumour cells interconnect to a functional and resistant network. *Nature* **528,** 93–98 (2015).

34. Jung, E. *et al.* Tumor cell plasticity, heterogeneity, and resistance in crucial microenvironmental niches in glioma. *Nat. Commun.* **12,** 1014 (2021).

35. Davare, M. A. *et al.* Rare but recurrent ROS1 fusions resulting from chromosome 6q22 microdeletions are targetable oncogenes in glioma. *Clin. Cancer Res.* **24,** 6471–6482 (2018).

36. Zhang, X., Lv, Q.-L., Huang, Y.-T., Zhang, L.-H. & Zhou, H.-H. Akt/FoxM1 signaling pathway-mediated upregulation of MYBL2 promotes progression of human glioma. *J Exp Clin Cancer Res* **36,** 105 (2017).

37. Yang, W. *et al.* Sex differences in GBM revealed by analysis of patient imaging, transcriptome, and survival data. *Sci. Transl. Med.* **11,** (2019).

38. Wang, S. *et al.* Integrated analysis of 34 microarray datasets reveals CBX3 as a diagnostic and prognostic biomarker in glioblastoma. *J. Transl. Med.* **17,** 179 (2019).

39. Kong, J. *et al.* Machine-based morphologic analysis of glioblastoma using whole-slide pathology images uncovers clinically relevant molecular correlates. *PLoS One* **8,** e81049 (2013).

40. Liberzon, A. *et al.* The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* **1,** 417–425 (2015).

41. Mao, P. *et al.* Mesenchymal glioma stem cells are maintained by activated glycolytic metabolism involving aldehyde dehydrogenase 1A3. *Proc. Natl. Acad. Sci. USA* **110,** 8644–8649 (2013).

42. Beckner, M. E., Pollack, I. F., Nordberg, M. L. & Hamilton, R. L. Glioblastomas with copy number gains in EGFR and RNF139 show increased expressions of carbonic anhydrase genes transformed by ENO1. *BBA Clin.* **5,** 1–15 (2016).