Functional diversity and cooperativity between subclonal populations of pediatric glioblastoma and diffuse intrinsic pontine glioma cells

Mara Vinci1,2,3, Anna Burford1,2,3, Valeria Molinari1,2,3, Ketty Kessler1,2,3, Sergey Popov1,2,3,4, Matthew Clarke1,2,3, Kathryn R. Taylor1,2,3,5, Helen N. Pemberton6, Christopher J. Lord6, Alice Gutteridge7, Tim Forshaw7, Diana Carvalho1,2,3, Lynley V. Marshall8, Elizabeth Y. Qin5, Wendy J. Ingram9,10, Andrew S. Moore9,10,11, Ho-Keung Ng12, Saoussen Trabelsi13,9,10, Andrew S. Moore

The failure to develop effective therapies for pediatric glioblastoma (pGBM) and diffuse intrinsic pontine glioma (DIPG) is in part due to their intrinsic heterogeneity. We aimed to quantitatively assess the extent to which this was present in these tumors through subclonal genomic analyses and to determine whether distinct tumor subpopulations may interact to promote tumorigenesis by generating subclonal patient-derived models in vitro and in vivo. Analysis of 142 sequenced tumors revealed multiple tumor subclones, spatially and temporally coexisting in a stable manner as observed by multiple sampling strategies. We isolated genotypically and phenotypically distinct subpopulations that we propose cooperate to enhance tumorigenicity and resistance to therapy. Inactivating mutations in the H4K20 histone methyltransferase KMT5B (SUV420H1), present in <1% of cells, abrogate DNA repair and confer increased invasion and migration on neighboring cells, in vitro and in vivo, through chemokine signaling and modulation of integrins. These data indicate that even rare tumor subpopulations may exert profound effects on tumorigenesis as a whole and may represent a new avenue for therapeutic development. Unraveling the mechanisms of subclonal diversity and communication in pGBM and DIPG will be an important step toward overcoming barriers to effective treatments.

GBM and DIPG are a highly heterogeneous group of high-grade glial tumors with no effective treatments.1 Integrated molecular profiling1–7 has revealed unique, specific and highly recurrent mutations in genes encoding histone H3 variants that mark robust subgroups of pGBM and DIPG with distinct age of onset, anatomical distribution, clinical outcome, and histopathological and radiological features.8–10. A paradigm shift away from extrapolating from inappropriate adult GBM data and toward a more pediatric-biology-specific approach to developing new therapies has been a positive consequence of the discovery of these mechanisms of tumorigenesis10–12.

Despite these advances in our understanding of the unique biological drivers of these diseases6, a major challenge to improving outcomes for children with these tumors is likely to overlap with morphologically similar tumors in adults: their extensive intratumoral heterogeneity11. This has been demonstrated spatially by the application of genomic analyses of topographically distinct areas of the tumor at resection13, through longitudinal studies of tumor...
progression and recurrence\textsuperscript{14}, and through single-cell RNA sequencing of bulk primary tumor specimens\textsuperscript{15}. All of these analyses suggest the presence of multiple coexisting tumor subclones that may be important to the proliferative and invasive capacities of the tumor, as well as cell fate decisions in response to the tumor microenvironment and selective pressure associated with therapeutic intervention. The relative contributions to the tumorigenic phenotype of these subclones is unclear, as is to what extent they interact during the tumor’s evolutionary history—key factors in understanding the implications for novel treatment strategies\textsuperscript{13}.

In adult GBM, multiple subclones may also be marked by differential, mutually exclusive gene amplification events present in an individual tumor\textsuperscript{22–24}, an observation also reported in isolated specimens of DIPG\textsuperscript{25,26}. In these examples, cells harboring distinct receptor tyrosine kinase gene amplifications were found intermingled throughout tumor specimens in a manner that suggested an environment conducive to the coexistence of multiple cellular subpopulations\textsuperscript{27–29}. Two-dimensional (2D) mapping of these subclones across specimens showed some evidence of a predilection of certain subclones for perivascular niches, invasive tumor fronts, or the periphery of necrotic areas\textsuperscript{28,29}. In evolutionary biology terms, this stable coexistence in conjunction with a degree of specialization appears to imply cooperativity\textsuperscript{4}. This posits a selective advantage for an interactive cellular network and promotes biological diversity within a tumor population as an important driver of the malignant phenotype in these cancers.

With pGBM and DIPG harboring considerably fewer somatic mutations than adult GBM\textsuperscript{1}, we sought to investigate the possibility of tumor heterogeneity reflecting cooperation of subclones in what we consider to be an ideal model system for cancers sharing these histologies. Through an integrated approach of single and multiple sequencing strategies of patient samples coupled with in vitro isolation of subclonal populations, we concluded that biological diversity is selected for across time and space, with genotypically and phenotypically distinct tumor compartments working together to enhance key tumorigenic features such as invasion and migration.

**Results**

**pGBM and DIPG comprise multiple subclones.** We reanalyzed whole-genome and exome sequencing from 142 recently published pGBM and DIPG specimens for which matched germline data were available\textsuperscript{27–29}. We calculated the cancer cell fractions (CCF) for all somatic single nucleotide variants (SNVs) and small insertions or deletions, taking into account the implied tumor cell percentage, overall ploidy, local copy number alterations and loss of heterozygosity\textsuperscript{27,28} (Supplementary Table 1). In almost all cases, we observed a complex inferred subclonal architecture suggestive not of a single clonal expansion, but of multiple codominant subclonal populations, regardless of tumor location (n = 93 DIPG, n = 20 other midline, n = 29 cerebral hemispheres) or histone mutation subgroup (n = 10 H3.3 G34R (H3F3A), n = 61 H3.3 K27M (H3F3A), n = 23 H3.1 K27M (HIST1H3B, HIST1H3C), n = 48 histone wild-type) (Fig. 1a). Despite this variability in the fraction of any tumor harboring a given mutation, at a gene level there were certain recurrent mutations that were found to be consistently clonal (H3F3A, HIST1H3B, HIST1H3C, ATRX, NF1), some that were found to be predominantly clonal, but with some subclonal examples (ACVR1, TP53), and some frequently found in subclonal populations (ATM, PIK3R1, PPM1D, PDGFR, BRAF, PIK3CA) (Fig. 1b). These data provide important evidence for the likely timing of these mutations during tumor evolution. Using the EXPANDS package\textsuperscript{27,28}, we used the sequencing data to predict an absolute number of subclones present in each tumor sample, deriving a median number of 6 (range 1–14), with more than 85% of tumors appearing to harbor 3–10 subclones (Fig. 1c and Supplementary Table 2). The percentage of clonal alterations ranged from 100% (n = 1) to 5.2% (median = 35.0%) (Supplementary Fig. 1a). There was a direct relationship between the overall mutational burden (number of somatic coding SNVs) and number of subclones (Pearson r² = 0.2188, P = 4.36 × 10⁻⁴), though with several outliers (Fig. 1d). There were no differences in subclonal number between different anatomical sites (Fig. 1e), despite the differing survival times by tumor location\textsuperscript{11}. pGBM with H3.3 G34R mutations had a significantly elevated number of subclones compared with other tumors (median = 8.5, P = 0.044, t-test), while there were significantly fewer in infants (<3 years at diagnosis; median = 4, P = 0.0188) (Fig. 1e). Plotting the number of subclones against hazard ratios for overall survival in a similar manner to that described in a pan-cancer analysis\textsuperscript{11}, we identified tumors harboring more than 10 subclones to have the worst prognosis (relative risk = 3.3) (Supplementary Fig. 1b).
Inferred heterogeneity

Gene-level clonality

Subclonal architecture

Mutational burden

Location

H3 subgroup

Subclonal diversity
Fig. 2 | DIPGs infiltrate the brain through branching evolution and genotypic convergence. **a**, Thirteen different tumor-harboring regions of HSJD-DIPG-010 were sampled post mortem, from within and outside the pons. Scale bars, 100 μm. **b**, Exome sequencing was carried out for all regions. CCFs plotted as a heat map for all variants found in at least one specimen, with anatomical location highlighted and color-coded. **c**, Phylogenetic trees were reconstructed using neighbor-joining algorithms based on the nested subpopulation phylogenies calculated as part of EXPANDS, with evident laterally directed evolution and early escape from the pons of tumor cells found in distinct anatomical sites. GL, germline. **d-f**, Eight different tumor-harboring regions of HSJD-DIPG-014 subjected to the same analysis. Scale bars, 100 μm. **g-i**, Eight different tumor-harboring regions of HSJD-DIPG-015 subjected to the same analysis. Scale bars, 100 μm.

Although patients with H3.3 G34R mutations had a better prognosis ($P=3.94 \times 10^{-4}$, log-rank test), tumors with $\geq 10$ subclones nonetheless showed a trend toward a shorter survival time ($P=0.068$, log-rank test) (Fig. 1f). In multivariate analysis including location, age and subgroup, only H3.3 K27M mutations ($P=0.000082$, Cox proportional hazards model) and a number of subclones greater...
than 10 (P = 0.0082) were independent predictors of shorter survival (Supplementary Fig. 1c).

The tumor cohort studied was heavily enriched in DIPG samples, and owing to the unresectability of these lesions, it comprised a mixture of pre-treatment biopsy samples and post-treatment autopsy samples\(^{3,4,5}\). We observed no systematic differences in subclonal architecture when comparing samples taken at these differing time points, regardless of diagnosis or histone mutation status (Supplementary Fig. 1d). We were able to assess this directly for eight cases for which paired pre- and post-treatment sequencing data were available. By plotting the change in major subclonal tumor proportion over time, we observed changes in the proportion of individual subpopulations in response to therapy and tumor evolution; in all cases, however, several significant populations remained unchanged, and both before and after treatment the tumor was inferred to harbor multiple subclones, suggesting either equivalent fitness of multiple subclones or pressures restricting the ability of any given clone to sweep to fixation (Supplementary Fig. 1e).

DIPG cells escape the pons early during tumor evolution. More direct evidence for the presence of multiple, genetically distinct subclones could be seen from sequencing 62 topographically distinct samples from 14 different patients (Supplementary Table 3). Comparing the CCFs from across a given tumor sample clearly demonstrated both the ubiquitous presence of presumed driver alterations (histone mutations, NFI) (Supplementary Fig. 2a) but also a range of mutations private to only one portion of the tumor. Of note, each distinct tumor region itself was inferred to harbor multiple subclones.

The collection of DIPG samples at autopsy represented a unique opportunity to evaluate the spatial heterogeneity of these tumors. In one case (QCTB-R091/R092), distinct low-grade and high-grade components were manually dissected and found to harbor key oncogenic mutations in one and not the other region (for example, PIK3CA H1047R in grade IV and not grade II), in addition to ubiquitous drivers such as ACVR1 (Supplementary Fig. 2b). It has previously been shown that these diffusely infiltrating lesions may be found outside the pons and spread throughout the central nervous system at the time of death\(^3\). Multi-sample sequencing strategies allowed us to again identify early driver events present throughout the tumor cells of an individual patient (H3F3A, HIST1H3B), as well as those occurring only at the point of escape of cells from the brainstem, such as mutations in WNK2, known to act in glioma cell migration and invasion\(^1\) (Supplementary Fig. 2c). Across multiple sites in multiple samples (Fig. 2, Supplementary Fig. 3 and Supplementary Table 3), mapping SNVs and copy number aberrations revealed branching evolutionary trajectories. This was particularly evident in the most extensively sampled cases (Fig. 2), where distinct branches highlighted the profound laterality of tumor evolution, while tumor cells found in midbrain, cerebellar and thalamic regions were seen to diverge early from the pontine mass. While the difference in mutational profiles may be a result of invasive cells cycling more slowly, the presence of convergent or parallel evolution in key oncogenic drivers such as PIK3CA, NFI1, MKI67, NOTCH1 and DMNT3A (Supplementary Fig. 3) strongly suggests a predominantly early evolutionary divergence of cells that subsequently migrated outside the pons.

In vitro isolation of genotypically and phenotypically distinct subclones. To determine whether the subclonal tumor cell populations present in PGBM and DIPG represent functionally distinct entities (rather than simply reflecting stochastic alterations occurring as a result of increasing genetic instability), we devised a methodology to isolate and expand single tumor cells under stem cell conditions, referred to as ‘stem-like’ cells, in both 2D\(^2\) and three-dimensional (3D) culture\(^4\) for further analysis (Fig. 3a). Using this approach, we identified three primary patient-derived H3.3 K27M mutant samples (two DIPG, one thalamic pGBM) (Supplementary Fig. 4) from a well-characterized panel of six cultures (Supplementary Fig. 5a,b) to readily form single-cell-derived colonies in both 2D and 3D, at rates varying between 7.5 and 20.8% of cells (Fig. 3b). Colonies isolated from SU-DIPG-V1 were identified using high-content image analysis (Fig. 3c,f) and displayed highly variable growth characteristics in vitro when grown as 3D neurospheres (Fig. 3d) and in 2D on laminin (Fig. 3g). When sequenced at high depth using a custom-designed targeted panel (Supplementary Table 4), in addition to ubiquitously shared mutations (for example, H3F3A, TP53), around half the colonies harbored a series of shared mutations not seen in the remainder (for example, PRR51, CHD3), while most were also found to contain a series of private events restricted to individual cell populations, including genes associated with cell shape and motility (FLNC, CTTN, RANGAP1) (Fig. 3e,h). Individual laminin-grown colonies with fast (A-D10), intermediate (A-B8) and slow (A-E6) growth rates (Fig. 3i) were seen to have significantly differing capacities for invasion into Matrigel (Fig. 3j) and migration on fibronectin (Fig. 3k) in vitro. Thus, individual tumor samples contain a dynamic diversity of overlapping genotypic and phenotypic populations in the stem-like cell compartment.

Rare tumor subclones can harbor pathogenic variants driving differing phenotypes. For HSJD-DIPG-007, we were able to utilize the ability to isolate these genetically and phenotypically distinct subclonal populations to investigate the role of individual genotypes without needing to artificially engineer the cells. We identified a single-cell-derived neurosphere colony (NS-F10) as harboring a private mutation in the histone H4 methyltransferase KMT5B (SUV420H1) (Fig. 4a), which was found to be present in the original bulk primary culture in only 2 of 678 reads (0.295%) (Fig. 4b). This mutation results in the acquisition of a stop codon at amino acid position 187 (R187*), predicted to truncate the protein. Examining published sequencing datasets, we identified another case of pGBM from Schwartzentruber et al.\(^4\), FGBP18, as harboring a subclonal R699* truncating mutation of KMT5B in 12.2% of reads (Fig. 4b), demonstrating this is not a unique observation. By digital droplet PCR, we confirmed that this mutation was present in 49.77% (8,060 of 16,196) of droplets from NS-F10 (assuming heterozygosity, this reflects presence in 99.64% cells), present in 0.48% (108 of 22,512) of reactions from the original culture, and absent (not significantly different from normal human astrocyte control; 1 of 18,484, 0.009%) from a ‘natural isogenic’ (confirmed by exome sequencing) counterpart subclone NS-F8 (Fig. 4c). The KMT5B mutant (NS-F10) and wild-type (NS-F8) subclones did not show appreciable differences from each other, or from the heterogeneous original bulk HSJD-DIPG-007 cells, in terms of morphology or immunophenotype (Supplementary Fig. 5c). The methyltransferase encoded by the gene is involved predominantly in dimethylation and, to a lesser extent, trimethylation\(^5\) of histone H4K20, and consequently by immunofluorescence we observed a reduction in H4K20me2 in NS-F10 compared to HSJD-DIPG-007 bulk cells and NS-F8 (Fig. 4d). An unbiased drug screen of all three colonies against 80 chemotherapeutic and targeted agents (Supplementary Fig. 6a–c and Supplementary Table 5) revealed significantly enhanced sensitivity to multiple chemotypes of PARP inhibitors of the KMT5B mutant NS-F10 compared to wild-type NS-F8 and HSJD-DIPG-007 bulk cells (10–30 fold difference for talazoparib, 50% cell survival concentration (SF\(_{50}\)) 1 μM vs. 11 μM and 31 μM, respectively; 4.5-fold difference for olaparib, SF\(_{50}\) 0.85 μM vs. 3.83 μM and 3.80 μM, NS-F10 vs. NS-F8 and HSJD-DIPG-007 bulk population ANOVA P < 0.001 in each case) (Fig. 4e). Notably, when subclones were cocultured, rather than a dilution effect dependent on the relative proportions, mixed cultures were insensitive as the heterogeneous bulk population (Supplementary Fig. 6d). Thus the mutation appears to confer a loss of function on these cells, presumably due to an abrogated DNA repair process associated with...
loss of H4K20me2 (but not H4K20me3 or total H4, Supplementary Fig. 6e) and recruitment of 53BP1.

Distinct infiltrative phenotypes of genotypically divergent DIPG subclones in vivo. RNA sequencing analysis of the subclones revealed elevated gene expression in NS-F10 cells of a range of genes associated with remodeling the extracellular matrix (Fig. 4f and Supplementary Table 6). These included the fibronectin receptors α5- and αv-integrin, with differential protein expression validated by immunofluorescence and immunohistochemistry (Fig. 4g). Although there was a slightly enhanced growth capability of the heterogeneous HSJD-DIPG-007 bulk cells (Fig. 4h), NS-F10 and NS-F8 subclones were similar to each other in growth, though we did observe significant differences in invasion into Matrigel (Fig. 4i) and migration on fibronectin (Fig. 4j), even after growth had been controlled for. This absence of α-integrin expression likely underlies the inability of NS-F8 to migrate on fibronectin; of note, NS-F8 neurospheres also harbored a significantly reduced migratory capacity compared with NS-F10 on a range of other substrates, including tenasin-C, laminin and Matrigel (Supplementary Fig. 6l). In all instances, the mixed population bulk HSJD-DIPG-007 cells were significantly more migratory than either subclone.

The KMT3B wild-type NS-F8 cells had significantly reduced invasive and migratory capacities, which could be reversed (unlike those of the KMT3B mutant NS-F10 cells) upon culture with conditioned medium from the HSJD-DIPG-007 bulk cells (Fig. 4k), suggesting the presence of expressed factors absent from the isolated NS-F8 cultures. These cells also differentially responded to the chemokine CXCL2 in terms of a significantly enhanced migration on fibronectin (Fig. 4k). We chose this CXC ligand because it was one of the most differentially expressed genes by RNA sequencing analysis in NS-F10 (and HSJD-DIPG-007 bulk) compared to NS-F8 (Fig. 4f). Thus we have a model whereby paracrine signaling between subclones underlies the cooperative interactions observed in mixed populations.

In line with in vitro data, phenotypic differences were also recapitulated in vivo, where both bulk cell populations and NS-F10 subclones formed diffusely infiltrating tumors within 23–24 weeks after orthotopic implantation in the pons of NOD-SCID mice, whereas NS-F8 lesions were substantially less infiltrative and conferred a lower tumor burden, even after 30–32 weeks, despite there being little difference in proliferative capacity and immunophenotype in the brains (Fig. 5a). NS-F8 tumor-bearing mice also had a longer survival (median = 205 d (NS-F8) vs. 141.5 d (NS-F10) and 169 d (bulk), P = 0.0236, log-rank test) (Fig. 5b). Tumors from heterogeneous bulk cells were confirmed by digital droplet PCR to harbor a low subclonal frequency of KMT3B R187* mutation (0.23%) (Fig. 5c), showing no significant selective pressure against the heterogeneous population. Thus even rare tumor cell subclonal populations may have different behaviors in vitro and in vivo of importance to key phenotypic features of DIPG that currently preclude effective treatments.

Notably, we observed similar results in a second model. A slower growing subclone of SU-DIPG-VI in vitro, A-E6, formed a less cellular tumor (Fig. 5d) and had an extended survival when grown orthotopically in vivo of more than 118 d longer than a rapidly proliferating, highly invasive subclone (A-D10) and 154 d longer than the unselected bulk culture (P = 0.037, log-rank test) (Fig. 5e).

DIPG subclones cooperate to enhance tumorigenic phenotypes. To explore the nature of these subclonal interactions, we differentially labeled and cocultured genotypically and phenotypically distinct subclones from two DIPG samples: A-E6 and A-D10 from SU-DIPG-VI (Fig. 6a–d) and NS-F8 and NS-F10 from HSJD-DIPG-007 (Fig. 6e–h). When cultured in equal proportions, having been replated as single neurospheres, although there was little difference in observed growth rates (Fig. 6a,c,e), there was a marked enhancement of invasion and migration conferred on the poorly motile subclones by coculture with their more invasive and migratory counterparts (Fig. 6b,d,f,g). In both models, cell labeling allowed us to demonstrate that this was not a simple dilution effect of the mixture, but that the specific subclones otherwise lacking a pronounced ability to invade into Matrigel (Fig. 6d) or migrate on fibronectin (Fig. 6h) had markedly enhanced phenotypes, clearly colocalized and moving in concert alongside their natural isogenic pairs (Supplementary Videos 1 and 2). In vivo, cocultured NS-F8 and NS-F10 were found to retain their mixed proportions and infiltrate more extensively throughout the central nervous system than NS-F8 alone (Supplementary Fig. 6g), conferring shorter survival on mice harboring these orthotopic tumors (P = 0.045, log-rank test) (Supplementary Fig. 6l). Thus we conclude that there exists an actively maintained cooperative network of subclones within DIPGs that depends on strongly positive interactions to elicit the highly aggressive clinical phenotypes seen in children with this incurable disease.

Fig. 3 | Isolation of genotypically and phenotypically diverse single stem-like cell-derived subclones of pediatric GBM and DIPG. a, Isolation of subclonal populations: disaggregation of heterogeneous mixtures of patient-derived tumor cells, flow sorting into single cells in 96-well plates, and colony formation as either 2D cultures, adherent on laminin, or 3D neurospheres, all under stem cell conditions. Individual subclonal colonies are subjected to high-throughput phenotypic analysis and targeted resequencing, and further cultured for detailed in vitro and in vivo mechanistic comparison with heterogeneous bulk populations. b, Percentage of single cells that formed colonies under 2D laminin and 3D neurosphere stem cell conditions are given for six pGBM and DIPG primary patient-derived cell cultures, labeled by anatomical location, histone H3 mutation subgroup (dark green, H3F3A; light green, H3F3A) and name of the cell line. Mid, midline. e, 3D neurosphere culture from single-cell-derived colonies from SU-DIPG-VI assessed by Celigo S imaging cytometer. d, Growth of single-cell-derived colonies over time, assessed as diameter of neurosphere, labeled and color-coded. e, Targeted sequencing contingency plot of somatic mutations common to all subclones (blue), shared among certain subclones (yellow) and private to individuals (red). f, 2D laminin culture from single-cell-derived colonies from SU-DIPG-VI assessed by Celigo S imaging cytometer. g, Growth of single-cell-derived colonies over time, assessed as diameter of neurosphere, with subclones taken for later analysis highlighted: A-D10 (fast, purple), A-B8 (intermediate, pink) and A-E6 (slow, violet). h, Targeted sequencing contingency plot of somatic mutations common to all subclones (blue), shared among certain subclones (yellow) and private to individuals (red). Gene names are colored to highlight private mutations in selected subclones or common to A-D10 and A-B8 (brown). i, Time course for growth of selected subclones replated and grown over 160 h, highlighting statistically significant differences among subclones and heterogeneous bulk cell populations of SU-DIPG-VI (blue). Representative images at 72 h are provided from the Celigo S cytometer, with tumor cells marked in green. Data derived and representative images taken from n = 3 independent experiments. Scale bars, 500 µm. j, Time course of invasion of cells into a Matrigel matrix over 72 h, either as percentage of the total area in the field of view covered by invading cells, or as a percentage of time zero. Representative images given at 72 h, with extent of tumor cell invasion marked in green. Data derived and representative images taken from n = 3 independent experiments. Scale bars, 500 µm. k, Time course of tumor cell migration onto Matrigel over 72 h, either as percentage of the total area of the well covered by migrating cells or as a percentage of that at time zero. Representative images given at 72 h, with extent of tumor cell migration marked in green. Data derived and representative images taken from n = 3 independent experiments. Scale bars, 500 µm. ANOVA. *P < 0.05. **P < 0.01. ***P < 0.001. All graphs show mean ± s.d.
Discussion

Widespread intratumoral heterogeneity in human cancer has become a prevalent theme in high-throughput sequence analysis of tumor specimens, with critically important implications for the success of therapeutic targeting\(^3\). Less attention has been given to the functional implications of this subclonal diversity and the interactions between distinct tumor subpopulations. Here we utilize pGBM and DIPG as cancer types with a relatively low mutational burden, yet a high degree of heterogeneity, to isolate these genotypically and phenotypically different compartments and provide evidence that subclonal diversity is selected for as a result of cooperative interactions that promote tumorigenesis.

Single-cell-derived colonies were established and expanded under stem cell culture conditions, though without marker preselection, in contrast to a recent approach\(^3\). Phenotypic differences were observed, in terms of morphology, growth, migration and invasion, that could be linked directly to concurrent genotypic differences in the subclones. These properties, first identified through high-content screening during initial expansion, were maintained upon repassing in short-term culture, indicating inherently fixed characteristics in markedly different tumor cell subpopulations. SNVs differing among single-cell-derived colonies could be found at low frequencies in the original tumor mass and thus were reflecting not an acquired artifact of the culture conditions but instead a propensity of genotypically distinct subclones to harbor stem-like properties, further evidenced through their tumorigenic capacity in vivo. It has previously been proposed that a branched Darwinian evolution model integrated with a hierarchy of multiple cancer stem cell populations may help explain the spatial and temporal characteristics of observed intratumoral heterogeneity\(^1\), with evidence provided in leukemia\(^2\) and solid tumors\(^2\).

In our models, the differing phenotypes of individual subclones were substantially less prominent than in heterogeneous unsorted primary cultures, with enhanced growth, invasion and migration properties of mixed populations of cells supporting the interpretation of sequencing analyses suggesting that subclonal diversity is selected for spatially\(^4,5\) and temporally\(^5\) in these tumors. The maintenance of such stable coexistence during tumor evolution implies a degree of cooperativity. In our example, we isolated ‘natural isogenic’ subclonal populations differing by a key loss-of-function mutation in an H4K20 methyltransferase, in which the more migratory mutant cells were able to confer such properties on their wild-type counterparts, seemingly at least in part through expression of key chemokines such as CXCL2. A similar concept of ‘cooperative invasion’ was first identified in melanoma\(^6\), whereby phenotypically distinct subpopulations of cells were found to comigrate, a phenomenon also observed in DIPG in our genotypically distinct cells. Likewise, a recent elegant paper making use of a lentivirally transduced triple-negative breast cancer cell line allowed reconstruction of an aggressive phenotype in vivo using only two cooperating subclones: those overexpressing IL-11 and VEGFD\(^7\). Such a mechanism obviates the need for clonal selection to drive tumorigenesis and predicts the maintenance of intratumoral heterogeneity we observe. Notably, the proportion of cells that harbor these more enhanced phenotypes may be low and may therefore remain unidentified in bulk tumor profiling studies, though remaining critical in tumor development and maintenance.

pGBM harboring H3.3 G34R or G34V mutations were found to harbor a higher mutational burden and a greater subclonal diversity than other tumor subgroups. Although the mechanisms are not known, this likely reflects an underlying DNA repair defect associated with the inability of the mutant H3K36 to be trimethylated, disrupting its important function in mismatch repair\(^7,8\). Despite
this, these tumors do not have the mutational burden of hypermutator cases with biallelic mismatch repair deficiency, for whom immune checkpoint inhibitors appear to offer an exciting new therapeutic option. It is not clear, therefore, that patients with H3.3 G34R or G34V mutations would benefit from a similar strategy. Unfortunately, no H3.3 G34R or G34V cultures were available for our study, and most of our functional work was focused on H3.3 K27M mutant DIPG samples, which were more amenable to single-cell-derived colony formation in our assay than other tumor genotypes (although it is not clear whether this reflects imperfect culture conditions for these subgroups). It has previously been shown that these diffusely infiltrating lesions may be found outside the pons and spread throughout the central nervous system at the time of death. Reconstructing phylogenies through sequencing of tumor cells spread throughout the brain at autopsy indicates an early escape of migratory cells from the pons,
before the rapid proliferative expansion occurring by the time of presentation and treatment. This has important implications for locally delivered therapies and reopens the debate concerning the initial use of whole-brain irradiation in children with DIPG.

The later acquisition of convergent mutations in genes controlling key signaling pathways associated with proliferation at these distant sites also underlies the challenges in preventing tumor recurrence and/or metastasis at anatomically distinct sites in the central nervous system.

In summary, these data demonstrate that pGBM and DIPG harbor a complex admixture of genotypically and phenotypically distinct stem-like cells driving a functionally based intratumoral...
heterogeneity. Understanding how the derived subclones interact and adapt to the tumor microenvironment, and to therapy, will be a key requirement for maximizing patient benefit from existing treatment options. Future strategies aimed at disrupting these interactions may represent a new therapeutic approach in these diseases.

Methods

Methods, including statements of data availability and any associated accession codes and references, are available at https://doi.org/10.1038/s41591-018-0086-7.

Received: 26 July 2016; Accepted: 3 May 2018; Published online: 2 July 2018

Fig. 6 | DIPG subclones cooperate to enhance tumorigenic phenotypes. Individual subclones of SU-DIPG-VI (a–d) and HSJD-DIPG-007 (e–h) were differentially labeled and cultured either as pure populations or mixed in equal ratios. a, Growth of cocultured (yellow) and monocultured E6 (green) and D10 (red) cells plated as single neurospheres after 96 h, measured as diameter of the sphere. Scale bars, 500 μm. b, Invasion of cocultured (yellow) and monocultured E6 (green) and D10 (red) cells into Matrigel over 168 h, with area assessed by ImageJ software from representative images provided from the Celigo S cytometer under phase contrast and fluorescence. Cocultures and D10 have significantly enhanced invasive capabilities compared to E6. Data derived and representative images taken from n = 3 independent experiments. Scale bars, 500 μm. c, Migration of mono- and cocultured E6 (green) and D10 (red) cells on Matrigel, assessed by the number of differentially labeled distant cells at 24 h, with representative images provided from the IncuCyte Zoom live-cell analysis system under phase contrast and fluorescence. Cells from individual subclones have enhanced migratory properties when cultured together compared to alone. Data derived and representative images taken from n = 3 independent experiments. Scale bars, 500 μm. d, Confocal microscopy analysis of invasion of mono- and cocultured E6 (green) and D10 (red) cells into Matrigel after 4 d, with nuclei stained with DAPI. Poorly motile E6 cells are found to invade further and in greater numbers alongside D10 cells than when cultured alone. Representative images taken from n = 3 independent experiments. Scale bars, 200 μm. e, Growth of cocultured (yellow) and monocultured NS-F8 (green) and NS-F10 (red) cells plated as single neurospheres after 96 h, measured as diameter of the sphere, with representative images provided from the Celigo S cytometer under phase contrast and fluorescence. Data derived and representative images taken from n = 3 independent experiments. Scale bars, 500 μm. f, Invasion of cocultured (yellow) and monocultured NS-F8 (green) and NS-F10 (red) cells into Matrigel over 72 h, with area assessed by ImageJ software from representative images provided from the Celigo S cytometer under phase contrast and fluorescence. Cells from individual subclones have enhanced migratory properties when cultured together compared to alone. Data derived and representative images taken from n = 3 independent experiments. Scale bars, 500 μm. g, Migration of mono- and cocultured NS-F8 (green) and NS-F10 (red) cells on fibronectin, assessed by the number of differentially labeled distant cells at 48 h, with representative images provided from the IncuCyte Zoom live-cell analysis system under phase contrast and fluorescence. Cells from NS-F8 have enhanced migratory properties when cultured with NS-F10 compared to alone. Data derived and representative images taken from n = 3 independent experiments. Scale bars, 500 μm. h, Confocal microscopy analysis of migration of mono- and cocultured NS-F8 (green) and NS-F10 (red) cells on fibronectin after 3 d, with nuclei stained with DAPI. Poorly motile NS-F8 cells are found to migrate further and in greater numbers alongside NS-F10 cells than when cultured alone. Representative images taken from n = 3 independent experiments. Scale bars, 200 μm. All comparisons carried out by ANOVA, **P < 0.01. ***P < 0.001. All graphs show mean ± s.d.

References

1. Jones, C., Perryman, L. & Hargrave, D. Paediatric and adult malignant glioma: close relatives or distant cousins. Nat. Rev. Clin. Oncol. 9, 400–413 (2012).
2. Buczko, P. et al. Genomic analysis of diffuse intrinsic pontine gliomas identifies three molecular subgroups and recurrent activating ACVR1 mutations. Nat. Genet. 46, 451–456 (2014).
3. Fontebasso, A. M. et al. Recurrent somatic mutations in ACVR1 in pediatric midline high-grade astrocytoma. Nat. Genet. 46, 462–466 (2014).
4. Schwartzentrum, J. et al. Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma. Nature 482, 226–231 (2012).
5. Taylor, K. R. et al. Recurrent activating ACVR1 mutations in diffuse intrinsic pontine glioma. Nat. Genet. 46, 457–461 (2014).
6. Wu, G. et al. Somatic histone H3 alterations in pediatric diffuse intrinsic pontine gliomas and non-brainstem glioblastomas. Nat. Genet. 44, 251–253 (2012).
7. Wu, G. et al. The genomic landscape of diffuse intrinsic pontine glioma and pediatric non-brainstem high-grade glioma. *Nat. Genet.* 46, 444–450 (2014).

8. Sturm, D. et al. Paediatric and adult glioblastoma: multiformal (epi)genomic culprits emerge. *Nat. Rev. Cancer* 14, 102–107 (2014).

9. Gajjar, A. et al. Pediatric brain tumors: innovative genomic information is transforming the diagnostic and clinical landscape. *J. Clin. Oncol.* 33, 2986–2998 (2015).

10. Funato, K., Major, T., Lewis, P. W., Allis, C. D. & Tabar, V. Use of human embryonic stem cells to model pediatric gliomas with H3.3.K27M histone mutation. *Science* 347, 1529–1533 (2014).

11. Grasso, C. S. et al. Functionally defined therapeutic targets in diffuse intrinsic pontine glioma. *Nat. Med.* 21, 555–559 (2015).

12. Hashizume, R. et al. Pharmacologic inhibition of histone demethylation as a key player in epigenetic regulation of genomic integrity. *Cancer Cell.* 29, 1560–1563 (2016).

13. Johnson, B. E. et al. Mutational analysis reveals the origin and therapy-driven evolution of recurrent glioma. *Science* 343, 189–193 (2014).

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

15. Mcgrana, N. & Swanton, C. Biological and therapeutic impact of intratumor heterogeneity in cancer evolution. *Cancer Cell.* 27, 15–26 (2015).

16. Little, S. E. et al. Receptor tyrosine kinase genes amplified in glioblastoma exhibit a mutual exclusivity in variable proportions reflective of individual tumor heterogeneity. *Cancer Res.* 72, 1614–1620 (2012).

17. Snuderl, M. et al. Mosaic amplification of multiple receptor tyrosine kinase genes in glioblastoma. *Cancer Cell.* 20, 810–817 (2011).

18. Szerlip, N. J. et al. Intratumoral heterogeneity of receptor tyrosine kinases EGFR and PDGFRα amplification in glioblastoma defines subpopulations with distinct growth factor response. *Proc. Natl. Acad. Sci. USA* 109, 3041–3046 (2012).

19. 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).

20. Puget, S. et al. Mesoencephal transition and PDGFRα amplification/mutation are key distinct oncogenic events in pediatric diffuse intrinsic pontine gliomas. *PLoS One* 7, e30313 (2012).

21. Nowak, M. A. Five rules for the evolution of cooperation. *Science* 314, 1560–1563 (2006).

22. Boll, N. et al. Heterogeneity of genomic evolution and mutational profiles in multiple myeloma. *Nat. Commun.* 5, 2997 (2014).

23. Van Loo, P. et al. Allele-specific copy number analysis of tumors. *Nat. Genet.* 4453–4458 (2011).

24. Nowak, M. A. Five rules for the evolution of cooperation. *Science* 314, 1560–1563 (2006).

25. Boll, N. et al. Heterogeneity of genomic evolution and mutational profiles in multiple myeloma. *Nat. Commun.* 5, 2997 (2014).

26. Van Loo, P. et al. Allele-specific copy number analysis of tumors. *Nat. Genet.* 4453–4458 (2011).

27. Andor, N., Harness, J. V., Müller, S., Mewes, H. W. & Petritsch, C. EXPANDS: multiple myeloma. *Nat. Genet.* 45, 105–113 (2013).

28. Mackay, A. et al. Integrated molecular meta-analysis of 1,000 pediatric diffuse intrinsic pontine gliomas. *Nat. Med.* 20, e1051 (2014).

29. Mackay, A. et al. Integrated molecular meta-analysis of 1,000 pediatric diffuse intrinsic pontine gliomas. *Nat. Med.* 20, e1051 (2014).

30. Caretti, V. et al. Subventricular spread of diffuse intrinsic pontine glioma. *Acta Neuropathol.* 128, 605–607 (2014).

31. Moniz, S. et al. Loss of WNK2 expression by promoter gene methylation occurs in adult gliomas and triggers Rac1-mediated tumour cell invasiveness. *Hum. Mol. Genet.* 22, 84–95 (2013).

32. Pollard, S. M. et al. Glioma stem cell lines expanded in adherent culture have diffuse intrinsic high-grade glioma: implications for diagnostic biopsy and targeted therapeutics. *Acta Neuropathol. Commun.* 4, 1 (2016).

33. Nikbakht, H. et al. Spatial and temporal homogeneity of driver mutations in diffuse intrinsic pontine glioma. *Nat. Commun.* 7, 11185 (2016).

34. Sullivan, R. et al. Characterizing temporal genomic heterogeneity in pediatric high-grade gliomas. *Acta Neuropathol. Commun.* 5, 78 (2017).

35. Chapman, A. et al. Heterogeneous tumor subpopulations cooperate to drive invasion. *Cell. Rep.* 8, 688–695 (2014).

36. Murasuy, A. et al. Non-cell-autonomous driving of tumor growth supports sub-clonal heterogeneity. *Nature* 514, 54–58 (2014).

37. Pui, C. C. et al. A histone H3K36 chromatin switch coordinates DNA double-strand break repair pathway choice. *Nat. Commun.* 5, 4091 (2014).

38. Pfister, S. X. et al. SETD2-dependent histone H3K36 trimethylation is required for homologous recombination repair and genome stability. *Cell Rep.* 7, 2006–2018 (2016).

39. Bouffet, E. et al. Immune checkpoint inhibition for hypermutant glioblastoma multiforme resulting from germline biallelic mismatch repair deficiency. *J. Clin. Oncol.* 34, 2206–2211 (2016).

40. Castel, D. et al. Histone H3F3A and Hist1H3B K27M mutations define two subgroups of diffuse intrinsic pontine gliomas with different prognosis and phenotypes. *Acta Neuropathol.* 130, 815–827 (2015).

Acknowledgements

We would like to thank L. Howell for confocal microscopy expertise, M. Lambros for technical advice, and H. Vogel for neuropathology expertise and assistance with autopsies at Stanford University. We are indebted to the multidisciplinary teams at the Royal Marsden Hospital, St George’s Hospital and Kings College Hospital for their continued assistance with prospective sample collection. Finally we thank the many children and families who contributed to this study through the donation of tumor tissue. This study makes use of data generated by the St Jude Children’s Research Hospital – Washington University Pediatric Cancer Genome Project, C. Hawkins and the Hospital for Sick Children, and the McGill University-DKEZ, Pediatic Brain Tumour Consortium. This work was supported by Cancer Research UK (grants CJ13468/A13982 and CJ13468/A25336), Abbvie’s Army and the DIPG Collaborative, and the INSTINCT network funded by The Brain Tumour Charity, Great Ormond Street Children’s Charity and Children with Cancer UK. We further acknowledge funding from the Brainchild Foundation (Australia), Children’s Hospital Foundation Queensland, the Xarba de Bancs de Tumors Acad. Sci. Catalunya (XBTC) sponsored by Pla Director d’Oncolegia de Catalunya, ISCIII, FEDER (CP13/00189), McKenna Claire Foundation, and the US National Institutes of Health (grants K08NS070926 and R01NS092597). We acknowledge the support of the Queensland Children’s Tumour Bank (QCTB) for provision of samples and clinical data. The QCTB is supported by the Children’s Hospital Foundation (Queensland) and the Brainchild Foundation. The authors acknowledge NHS funding to the NIHR Biomedical Research Centre at The Royal Marsden and the ICR and Experimental Cancer Medicines Centre (ECMC) funding.

Author contributions

M.V., A.B., A.M. and C.J. conceived the study and wrote the manuscript with contributions from all authors. M.V., A.B., Y.M., K.K., S.P., M.C., K.R.T., H.N.P., C.J.L., A.B., V.M., K.K., S.P., M.C., K.R.T., H.N.P., C.J.L., A.G., T.F. and D.C. carried out experiments. M.V., A.B., A.M. and C.J. analyzed data. L.V.M., E.Y.Q, W.J.I., A.S.M., H.-K.N., S.T., D.H.-B.B., N.E.-W., S.V., H.C.M., L.R.B., A.J.M., S.A.-S., C.C., J.M., C.d.T., O.C., M.S., A.M.C., M.M. and A.M. prepared samples and provided clinical annotations. All authors approved the manuscript.
Methods
Published sequencing data. Raw data were obtained from the European Genome-phenome Archive (https://www.ebi.ac.uk/ega/home) from five published sequencing studies and provided under data access agreements from the St. Jude Children's Research Hospital—Washington University Pediatric Cancer Genome Project (accession code EGAS00001000192)2 and The Hospital for Sick Children (EGAS00001000575)3 and the McGill University–DKFZ Pediatric Brain Tumour Consortium (EGAS00001000226)4 and EGAS00001000720)5. We also included data from our own study (EGAS00001000572)6 and from four tumors collected via the Institute of Cancer Research (South West London MREC-approved study 10/H0803/126 with full consent) included in a recent International Cancer Genome Consortium (ICGC) study (EGAS00001001139)7, all of which were also part of a recent genomics meta-analysis by our group8, processed data from which are housed at https://pediportoal.org/. In total, we obtained whole-genome (n=70) or exome (n=72) data from 142 pGBM and DIPG patients for whom matched normal DNA was available, six of which also had data from paired longitudinal sampling. The median age was 6.8 years at diagnosis and the median survival 11.45 months (Supplementary Table 2).

Patients and samples. All patient material studied under South West London Research Ethics Committee approval. We obtained longitudinal paired samples from two patients from the Centre Hospitalier Régional et Universitaire Hautepietre, Strasbourg, France; five DIPG patients with multiple sampling taken at autopsy from the Hospital San Joan de Deu, Barcelona, Spain; five DIPG patients with multiple sampling taken at autopsy from Stanford Medical School, Stanford, CA, USA; three patients with multiple samples from the Queensland Children's Tumour Bank, Brisbane, Australia; and one patient each with whole samples from St Georges Hospital and Kings College Hospital, London, UK (Supplementary Table 3), all of which were collected locally after informed consent. The four previously sequenced patient samples were obtained from the Chinese University of Hong Kong, China (n=3) and University Hospital Sousse, Tunisia, n=1). DIPG DNA was extracted from formalin-fixed, paraffin-embedded tissue following the DNeasy Blood & Tissue kit protocol (Qiagen, Crawley, UK). DNA was extracted from FFPE material from either 20-μm ribbons (n=2–4 per sample) or 5-μm sections cut onto slides (n=10 per sample). Slides were hydrated through an ethanol series before manual microdissection into a tube using a sterile fine needle. All tissue was incubated overnight with proteinase K at 56 °C with a further 1 h at 95 °C. DNA was extracted using the QIAamp DNA FFPE tissue kit protocol (Qiagen, Crawley, UK) using 360 μL of Buffer AL and 360 μL of ethanol, and eluted using 25 μL of 10 M Tris buffer at pH 8.5 for 7 min. Matched normal DNA was extracted from blood samples using the DNeasy Blood & Tissue kit (Qiagen, Crawley, UK). Concentrations were measured using a Qubit fluorometer (Life Technologies, Paisley, UK), with at least 400 ng sent for exome sequencing at the Tumour Profiling Unit, ICR, London, UK using the 50 Mb Agilent SureSelect platform (Agilent, Santa Clara, CA, USA), and paired-end sequenced on an Illumina HiSeq2000 (Illumina, San Diego, CA, USA) with a 100-bp read length. The average median coverage was 148× for the tumor exomes and 108× for tumor genomes.

Sequence analysis. For both published and newly generated raw sequencing data, reads were aligned to the hg19 build of the human genome using bwa v0.7.5a (http://bio-bwa.sourceforge.net/) and PCR duplicates removed with PicardTools 1.5 (https://broadinstitute.github.io/picard/). Somatic single nucleotide variants were called using the Genome Analysis Tool Kit v3.3-0 based on current best practices using local realignment around insertions or deletions, down-sampling for tumor genomes. Structural variants were called from whole-genome data using BreakDancer (http://breakdancer.sourceforge.net/) filtered to remove commonly multi-mapped regions to identify somatic breakpoints separated by a minimum of 10 kbp involving at least one Ensembl gene. Variants were annotated using the Ensembl Variant Effect Predictor v71 (https://wwwensembl.org/info/docs/tools/vep/) incorporating SIFT (http://sift.jcvi.org/) and PolyPhen (http://genetics.bwh.harvard.edu/pph2/) predictions, COSMIC v64 (http://www.sanger.ac.uk/genetics/CGP/cosmic/) and dbSNP build 137 (http://www.ncbi.nlm.nih.gov/sites/SNP) annotations. Somatic variants used for further subclonal analysis (non-synonymous, exonic or splicing) were genotyped using at least read 10 reads in both tumor and normal sequences. Copy number was obtained by calculating log, ratios of tumor/normal coverage binned into exons of known genes, smoothed using circular binary segmentation (https://broadinstitute.org/packages/release/bioc/html/DNAscopy.html) and processed using in-house scripts. To infer the proportion of the sample carrying each somatic variant in each sample, we calculated the cancer cell fraction (CCF) for each somatic variant9. Briefly, we determined the somatic allele-specific copy number profiles using read depth from whole-genome or exome sequencing as above analyzed by ASCAT10, which also provided for an estimate of the non-neoplastic cell contamination of the sample as well as the overall ploidy of the tumor. Loss of heterozygosity (LOH) was also calculated using ASCAT. Slides were analyzed on a microarray scanner (Affymetrix GeneChip 3000-7G), and the copy number, LOH and tumor cell purity were then used to calculate the CCF, which estimates the percentage of tumor cells carrying each mutation9, and truncated to 100% where experimental variability in sequence reads produced a value greater than this figure. Intratumoral heterogeneity and the number and frequency of subpopulations within individual tumor samples were calculated with the EXPANDS algorithm using evolutionary biology principles including the Shannon and Simpson indices and allowing for the concept that subclones may share a subset of variants that may be nested within each other10. This used copy-number-corrected variant allele frequencies of all somatic coding mutations clustered based on their cell-frequency probability distributions, and subject to pruning to assign individual mutations to predicted subpopulations11. For multi-region samples from the same patient, distance matrices derived from the cancer cell fractions of non-synonymous somatic coding mutations in each sample were used to construct phylogenies based upon neighbor-joining algorithms utilizing the nested subpopulation calculated as part of EXPANDS, and visualized using the ape package (v3.1-4) in R. For paired longitudinal samples taken pre- and post-treatment, we fitted a kernel density estimate of cell-frequencies at both time points and identified congegating clusters using a heat map visualization of the resulting biplot11. A customized R function identified the x and y coordinates of each cluster centroid, which served as an estimate of the number and relative composition of major subclones present in each sample. These were plotted on- and post-treatment with colored lines highlighting the inferred relationship between each cluster.

Cell culture. pGBM and DIPG patient-derived cultures were established either immediately after collection (biopsy, resection or autopsy) or from cryopreserved tissue, with authenticity verified using short tandem repeat (STR) DNA fingerprinting and certified mpoxless-free. SU-DIPG-IV and SU-DIPG-V (Milenyi Biotec Ltd. Bisley, UK) and HeLa (2ng/mL) (Stem Cell Technologies, Vancouver, BC, Canada). Thalamic H3.3 K27M pGBM QCTB-R059 and DIPGs (3D) neurospheres12. Cortical pGBM cultures ICR-G358 and HSDJ-GBM-01 were cultured in a serum-free medium composed of the neural stem cell medium RHB-A (StemCells, Inc. Cambridge, UK) supplemented with human bFGF (20 ng/mL), human EGF (20 ng/mL), human PDGF-AB (20 ng/mL) (Milenyi Biotec Ltd. Bisley, UK) and Heparin (2ng/mL) (Stem Cell Technologies, Vancouver, BC, Canada). Thalamic H3.3 K27M pGBM QCTB-R059 and DIPGs HSJ-DIPG-007, SU-DIPG-IV and SU-DIPG-V were cultured in a serum-free medium designated as Tumor Stem Medium (TSM) as previously described12, consisting of 1:1 Neurobasal-A (Invitrogen, Carlsbad, CA), and DMEM-F12 (Life Technologies) supplemented with HEPES, NEAA, Glutamaxx, sodium pyruvate (Life Technologies) and B27 (A) (Invitrogen, Carlsbad, CA), human bFGF (20 ng/mL), human EGF (20 ng/mL), human PDGF-AB (20 ng/mL), human PDGF-BB (10 ng/mL) (Shenandoah, BioTech, Warwick, PA) and heparin (2ng/mL) (Stem Cell Technologies, Vancouver, BC, Canada).

Establishment of single-cell colonies. Primary cultures were single-cell flow sorted into the inner 60 wells of a 96-well cell plate using a FACSAria I instrument (BD) equipped with an automated cell deconvolutor unit. Single cells were dropped in 100 μL per well of the same medium as described above, with the addition of penicillin and streptomycin (Life Technologies). The outer 16 wells were filled with 200 μL per well of PBS to avoid evaporation of medium. 96-well plates were incubated at 37 °C, 5% CO2, 95% humidity, and cells refed twice weekly with 10–20 μL of medium per well. Fully automated image analysis of single-cell-derived colonies in 2D and 3D was carried out on a Celigo S cytometer (Nexcelom Inc.)53. At indicated time points, 96-well plates were scanned, images acquired and growth assessed using the Confluence application for 2D adherent culture on laminin and the Tumourosphere application for determining the diameter of the neurospheres. Single-cell-derived adherent colonies were collected when they reached approximately 80% confluence, while the neurospheres were collected at around 700–800 μm diameter. On collection day, 10% of the cells were used to expand individual subclonal cultures, with the remaining 90% used for DNA extraction after overnight incubation with Proteinase K and RNase A using the QIamp DNA micro kit (Qiagen) and elution using 25 μL of 10 mM Tris buffer pH 8.5 for 5 min before quantification. A minimum of 50 ng DNA was used for targeted resequencing using a custom Agilent SureSelect panel of 455 genes recurrently mutated in pGBM or DIPG, or including all members of the histone gene family (Supplementary Table 4).

High-throughput assays and content image analysis. 3D invasion assays were performed as previously described13,14, with some modifications. Briefly, a total of 100 μL medium was removed from ULA 96-well round-bottomed plates containing neurospheres 250–300 μm in diameter (given the different growth rate among the bulk cell, subclone, single-cell-derived, and double-diffusion colony, cells were sorted on 2D and 300 μm in diameter to obtain similar-size neurospheres). Matrigel (100 μL) was gently added to each well (6 replicates) and plates were incubated at 37 °C, 5% CO2, 95% humidity for 1h.
Once the Matrigel solidified, 100 µL per well of culture medium was added on top. Starting from time zero, and at intervals up to 72 h, automated image analysis was carried out on a Celigo S imaging cytometer using the Confluence application. The degree of cell spreading in the Matrigel was measured and the data plotted either as percentage of total area in the field of view covered by invading cells or as percentage of initial size of each neurosphere at time zero (t = 3). 3D migration assays were similarly performed as previously described, with some modifications. Briefly, flat-bottomed 96-well plates (Greiner Bio-one) were coated for 2 h at room temperature with 50 µL per well of fibronectin, laminin, tenasin (Sigma–Aldrich) 10 µg/mL in PBS with calcium and magnesium, or 125 µg/mL Matrigel (Corning) in culture medium in absence of growth factors. Once coating was completed, a total of 200 µL per well of culture medium was added. For stimulation assays, cells were exposed to 100 ng/mL of EGF (BD Biosciences), 10 ng/mL FGF-2 (Santa Cruz Biotechnologies) or 20 ng/mL NGF (Millipore) in the presence of 2% FBS (w/v). Two percent of the assay volume was removed from ULA 96-well round-bottomed plates containing neurospheres 250–300 µm in diameter, and the remaining medium including the neurospheres were transferred to the precoated plates. Starting from time zero, and at intervals up to 72 h, automated image analysis was carried out on a Celigo S cytometer using the Confluence application. The degree of cell spreading on the different matrices was measured and data plotted either as percentage of total area in the well covered by migrating cells or as percentage of the initial size of each neurosphere at time zero (t = 3).

Digital droplet PCR. Digital droplet PCR was carried out on genomic DNA extracted from normal human astrocytes, heterogeneous HSJD-DIPG-007 bulk cells and subclones NS-F10 and NS-F8 using primers designed to detect KMT5B R187* (forward: GCCAATATTCCAATCCAGTCTCAGT; reverse: GCCAGGATATTCAATATTAAAGTTATTCTTCTT). On a QX200 digital PCR platform (Bio-Rad). Reporter sequences were CAAACATTGCAATAAAT (VIC, wild-type), CAGAATTTTTTTTTTTTTTTTTT (FITC, mutant). Each plate of 20-µL droplets consisted of 10 µL dfPCR Supermix for Probes (no dUTP, Bio-Rad), primers and probes at the same molar concentrations as used in qPCR, DNA up to 50 ng, and molecular biology grade water. Each reaction was homogenized and partitioned into a theoretical maximum of around 23,000 droplets by creating an emulsion with Droplet Generation Oil for Probes (Bio-Rad). The 0.85-µL droplets were then amplified using standard PCR cycling parameters and an annealing temperature of 60 °C in accordance with the manufacturer’s recommendations. At endpoint, the fluorescence of each individual droplet was read on the droplet reader to identify presence or absence of mutant and wild-type target sequences. The QuantSoft program (v1.4) fitted the droplet counts to a Poisson distribution to enumerate the DNA copies, from which the DNA concentration and mutant fraction could be calculated.

Drug screening. An in-house drug library consisting of 80 drugs used either in clinical practice or in late-stage development was screened. Each compound was dissolved in 100% dimethyl sulfoxide (DMSO) to give 5 mM stocks and then diluted to 0.5, 0.05, 0.005 and 0.0005 mM stocks in 96-well two-dimensional matrix plates. Daughter plates in 384-well format were prepared from these 96-well two-dimensional matrix plates. Plates were incubated in complete medium, counted and seeded into 96-well round-bottom ULA plates (Corning) at 1,000 cells per well, either in monoculture or in coculture (50:50), and allowed to form a single neurosphere per well (n = 6). One or 2 d after seeding, migration assays were performed on fibronectin, or Matrigel-D10 and E6) coated 96-well flat-bottom plates (Essen Bioscience), and brightfield and fluorescence images were acquired on an Incucyte ZOOM (Essen Bioscience). Images of a region of interest of identical size across all replicates (n = 6) were imported into ImageJ software and the number of cells migrated (at 12 h for D10 and E6 and 48 h for NS-F10 and NS-F8) was manually counted using the cell counter plugin and normalized to the cell ratio (100% for the monolayers and 50% for the cocultures). Growth was assessed using the Celigo 5 as above, while invasion was measured as area covered using ImageJ upon image calibration using a 1-mm graticule. Time lapse videos were also acquired using a Zeiss LSM700 confocal microscope with images acquired every 30 min.

In vivo orthotopic xenograft. All experiments were performed after approval by the Animal Welfare and Ethical Review Board at Institute of Cancer Research, in accordance with the UK Home Office Animals (Scientific Procedures) Act 1986, the United Kingdom National Cancer Research Institute guidelines for the welfare of animals in cancer research and the ARRIVE (Animal Research: Reporting In Vivo Experiments) guidelines. A single-cell suspension from heterogeneous bulk cells or subclones (HSJD-DIPG-007, NS-F10, NS-F8 or coculture, in Matrigel; SU-DPG-VI, A-D10 and A-E8, in medium) was prepared immediately before implantation in four to eight nude SCID (HSJD-DIPG-007 and subclones) or nude (Ncrs Foxn1nu) mice (SU-DPG-VI and subclones) randomly allocated per group at P35. Animals were anaesthetized with ketamine and xylazine (100 mg/kg and 5 mg/kg) and maintained under 1% isoflurane. The cranium was exposed via midline incision under aseptic conditions and 1 × 1 mm deep hole was drilled through the skull to the dura. Mice were placed in a stereotactic apparatus and 200,000 cells in 5 µL of B27-supplemented media were stereotactically implanted into the pons using a digital pump at an infusion rate of 2 µL/min and a 31-gauge Hamilton syringe. Coordinates used were 1.0 mm lateral to midline, 0.8 mm posterior to lambda, and −4 mm deep to cranial surface. At the completion of infusion, the syringe needle was withdrawn to minimize backflow of the injected cell suspension. Mice were then back to normal daily routine and inspected daily with no treatment. Mice were killed by CO2 asphyxiation at any time point, and brains were harvested for further analysis. Brain tissues were dissected at the level of the injection site, were embedded into paraffin, and were sectioned at 5 µm. Sections were stained with haematoxylin and eosin.

Human neuroblastoma cell line IMR-32C was purchased from the American Type Culture Collection (ATCC; Manassas, VA, USA) and cultured in minimum essential medium (MEM) supplemented with 10% fetal bovine serum (FBS). Cells were maintained in 75 cm2 tissue culture flasks. Cell on chamber slides were fixed with 4% paraformaldehyde at room temperature for 10 min and washed three times with phosphate-buffered saline (PBS) solution. Neurons were collected into conical tubes, centrifuged for 10 min at 146 × g, washed once with PBS and, after fixation (4% paraformaldehyde overnight 4 °C) and two washes in PBS and embedding into a microtome at 4 µm thickness. Cells were permeabilized with 0.5% Triton X-100 solution for 10 min at room temperature and then blocked with appropriate serum according to the species of secondary antibody for 1 h at room temperature. Secondary antibodies used were goat-anti mouse (A11010, ThermoFisher), donkey-anti rabbit (A11008, ThermoFisher). Primary antibodies directed against nestin (MAB3326 clone 10C2, Millipore, 1:100), SOX2 (MAB3326 clone 11-5B, Millipore, 1:200), TUN1 (MMS-435P, Covance, 1:2000), Olig-2 (Ab9610, Millipore, 1:200) and Musashi-1 (Ab5977, Covance, 1:2000) were washed and incubated overnight at 4 °C. Cells were then washed in PBS three times and incubated with Alexa Fluor 488- or Alexa Fluor 555-conjugated secondary antibodies 1 h at room temperature. For anti-HK327m3 (9733, Cell Signaling, 1:100) and anti-α-integrin (ab15031, Abcam, 1:100), samples were incubated at 37 °C for 20 min following by a secondary antibody incubation at 37 °C for 20 min. Nuclei were counterstained with DAPI and samples mounted with CellSaver (Vector Laboratories) and examined using a Leica DM2500 fluorescence microscope or a Zeiss LSM700 confocal microscope.

Coculture experiments. Laminin-adenherent cultures and neurospheres were dissociated and filtered through a 40-µm cell strainer to remove residual clumps. Single-cell suspensions were then incubated with CellTracker Green CMFDA (E6 and NS-F8 (Life Technologies) at a final concentration of 5 µM following the manufacturer’s instructions for suspension culture. Control unseeded cultures were incubated with an equivalent amount of DMSO. Once the staining protocol was completed, cells were washed once in complete medium, counted and seeded into 96-well round-bottom ULA plates (Corning) at 1,000 cells per well, either in monoculture or in coculture (50:50) and allowed to form a single neurosphere per well (n = 6). One or 2 d after seeding, migration assays were performed on fibronectin, or Matrigel-D10 and E6) coated 96-well flat-bottom plates (Essen Bioscience), and brightfield and fluorescence images were acquired on an Incucyte ZOOM (Essen Bioscience). Images of a region of interest of identical size across all replicates (n = 6) were imported into ImageJ software and the number of cells migrated (at 12 h for D10 and E6 and 48 h for NS-F10 and NS-F8) was manually counted using the cell counter plugin and normalized to the cell ratio (100% for the monolayers and 50% for the cocultures). Growth was assessed using the Celigo 5 as above, while invasion was measured as area covered using ImageJ upon image calibration using a 1-mm graticule. Time lapse videos were also acquired using a Zeiss LSM700 confocal microscope with images acquired every 30 min.
and Ki67 (M7240, Dako, 1:100). All primary antibodies were diluted into 1% Tris buffer solution with 0.05% Tween-20 except the Ki67 antibody, which was diluted into Dako antibody diluent, and staining was performed using an autostainer. Anti-human GFAP was incubated for 30 min and anti-H3K27me3 and anti-HNA for 1 h, all at room temperature. An Envision detection system (Dako K5007) was used for Ki67 staining, whereas for the others a Novocastra Novolink Polymer Detection Systems Kit (Leica Biosystem RE-7150) was used. Slides were then mounted using Leica CV Ultra mounting medium and assessed by an experienced pathologist (S.P.) blinded to cell identity.

**Statistical analyses.** Statistical analysis was carried out using R 3.3.0 (http://www.r-project.org/) and GraphPad Prism 7. Comparisons between groups of continuous variables employed Student’s t-test or the analysis of variance (ANOVA) test. Univariate differences in survival were analyzed by the Kaplan–Meier method and significance determined by the log-rank test. Multivariate analyses were carried out using the Cox proportional hazards model. All analyses were two-sided, and \( P < 0.05 \) after multiple testing correction was considered significant.

**Reporting Summary.** Further information on experimental design is available in the Nature Research Reporting Summary linked to this article.

**Code availability.** All custom scripts for data processing are available upon reasonable request.

**Data availability.** All new sequencing data are deposited in the European Genome-phenome Archive (https://www.ebi.ac.uk/ega/home) under accession code EGAS00001001436.

**References**

51. International Cancer Genome Consortium PedBrain Tumor Project. Recurrent MET fusion genes represent a drug target in pediatric glioblastoma. *Nat. Med.* **22**, 1314–1320 (2016).
52. Ding, L. et al. Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature* **481**, 506–510 (2012).
53. Vinci, M. et al. Advances in establishment and analysis of three-dimensional tumor spheroid-based functional assays for target validation and drug evaluation. *BMC Biol.* **10**, 29 (2012).
54. Vinci, M., Box, C. & Eccles, S. A. Three-dimensional (3D) tumor spheroid invasion assay. *J. Vis. Exp.* e52686, https://doi.org/10.3791/52686 (2015).
55. Vinci, M., Box, C., Zimmermann, M. & Eccles, S. A. Tumor spheroid-based migration assays for evaluation of therapeutic agents. *Methods Mol. Biol.* **986**, 253–266 (2013).
56. Brough, R. et al. Functional viability profiles of breast cancer. *Cancer Discov.* **1**, 260–273 (2011).
57. Postel-Vinay, S. et al. A high-throughput screen identifies PARP1/2 inhibitors as a potential therapy for ERCC1-deficient non-small cell lung cancer. *Oncogene* **32**, 5377–5387 (2013).
Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

### Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

- **n/a** Confirmed
- The **exact sample size** ($n$) for each experimental group/condition, given as a discrete number and unit of measurement
- An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
  - *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistics including **central tendency** (e.g. means) or other basic estimates (e.g. regression coefficient) AND **variation** (e.g. standard deviation) or associated **estimates of uncertainty** (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. $F$, $t$, $r$) with confidence intervals, effect sizes, degrees of freedom and $P$ value noted
  - *Give $P$ values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's $d$, Pearson's $r$), indicating how they were calculated
- Clearly defined error bars
  - *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

### Software and code

Policy information about availability of computer code

**Data collection**

For acquisition of quantitative in vitro cell culture data, the Confluence (2D) and Tumorsphere (3D) applications on the Celigo S imaging cytometer (Nexcelom Inc) were used. The QuantaSoft program (v1.4) aligned to the QX200 digital PCR platform (Bio-Rad) fitted the droplet counts to the Poisson distribution in order to enumerate the DNA copies, from which the DNA concentration and mutant fraction could be calculated.

**Data analysis**

Custom scripts for data handling are available upon request. For both published and newly generated raw sequencing data, reads were aligned to the hg19 build of the human genome using bwa v0.7.5a (bio-bwa.sourceforge.net), and PCR duplicates removed with PicardTools 1.5 (picard.sourceforge.net). Somatic single nucleotide variants were called using the Genome Analysis Tool Kit v3.3-0 based upon current Best Practices using local re-alignment around InDels, downsampling and base recalibration with variants called by the Unified Genotyper (www.broadinstitute.org/gatk/). Structural variants were called from whole genome data using Breakdancer (http://breakdancer.sourceforge.net) filtered to remove commonly multimapped regions to identify somatic breakpoints separated by a minimum of 10kb involving at least one Ensembl gene. Variants were annotated using the Ensembl Variant Effect Predictor v71 (www.ensembl.org/info/docs/variation/vep) incorporating SIFT (sift.jcvi.org) and PolyPhen (genetics.bwh.harvard.edu/pph2) predictions, COSMIC v64 (www.sanger.ac.uk/genetics/CGP/cosmic/) and dbsNP build 137 (www.ncbi.nlm.nih.gov/sites/SNP) annotations. Copy number was obtained by calculating log2 ratios of tumour/normal coverage binned into exons of known genes, smoothed using circular binary segmentation (www.bioconductor.org). Cancer cell fractions were calculated from somatic allele-specific copy number profiles derived by ASCAT (https://github.com/Crick-CancerGenomics/ascat). Intratumoral heterogeneity and the number and frequency of subpopulations within individual tumour samples were calculated using the EXPANDS algorithm (http://dna-discovery.stanford.edu/software/).
Distance matrices derived from the cancer cell fractions of non-synonymous somatic coding mutations in each sample were used to construct phylogenies based upon neighbour-joining algorithms utilising the nested subpopulation calculated as part of EXPANDS, and visualized using the ape package (v3.1-4) in R. RNA sequences were aligned to hg19 and organised into de novo spliced alignments using bowtie2 and TopHat2 (https://ccb.jhu.edu/software/tophat). Raw read counts and fragments per kilobase per million reads mapped (FPKPM) were calculated for all known Ensembl genes in assembly v74 using bedtools (http://bedtools.readthedocs.org) and Cufflinks (http://cole-trapnell-lab.github.io/cufflinks/).

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences
- Behavioural & social sciences
- Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

**Sample size**
For the genomic analyses, 142 pGBM and DIPG patients for whom matched germline data was available represented the sum total of publicly available data at the time. For the in vivo experiment, sample size was in part determined pragmatically as the maximum number that could be implanted in one day across all cell preparations to be tested, and was powered to detect a 20% difference in survival at an alpha of 0.05 and a power of 0.75.

**Data exclusions**
There were no data exclusions.

**Replication**
All in vitro experiments were carried out independently in triplicate.

**Randomization**
For the in vivo experiment, mice were randomly allocated to each individual cell preparation for implantation.

**Blinding**
Assessment of biomarkers in xenograft experiment carried out by an experienced pathologist blinded to cell identity

Reporting for specific materials, systems and methods

| Materials & experimental systems | Methods |
|---------------------------------|---------|
| n/a                             | n/a     |
| Involved in the study           | Involved in the study |
| Unique biological materials     | ChIP-seq |
| Antibodies                      | Flow cytometry |
| Eukaryotic cell lines           | MRI-based neuroimaging |
| Palaeontology                   |         |
| Animals and other organisms     |         |
| Human research participants     |         |

**Antibodies**

Primary antibodies directed against nestin (MABS326 clone 10C2, Millipore, 1:400), SOX2 (3579, Cell Signalling, 1:400), GFAP
(Z334, Dako, 1:50), CNPase (MAB326 clone 11-58, Millipore, 1:200), Tuj-1 (MMS-435P, Covance, 1:2000), Olig-2 (Ab9610, Millipore, 1:200), and Musashi-1 (Ab5977, Millipore, 1:200), anti-H3K27me3, (9733, Cell Signalling, 1:100) and anti-alpha-5 integrin (ab15031, Abcam, 1:100), human GFAP (M0761 clone 6F2, Dako, 1:300), and Ki67 (M7240, DAKO, 1:100) were used. Secondary antibodies used were goat anti-mouse (A11001, ThermoFisher), donkey anti-rabbit (A11008, ThermoFisher) and goat anti-rabbit (A11008, ThermoFisher).

**Validation**

Validated using appropriate positive control tissue as recommended by the manufacturer, as well as no primary antibody negative controls.

### Eukaryotic cell lines

**Policy information about cell lines**

| Cell line source(s)   | Patient-derived primary cultures developed by the authors. |
|-----------------------|-----------------------------------------------------------|
| Authentication        | STR profiling, custom panel and exome sequencing.        |
| Mycoplasma contamination | All were verified mycoplasma free.                      |
| Commonly misidentified lines (See [ICLAC register](#)) | No commonly misidentified lines were used. |

### Animals and other organisms

**Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research**

| Laboratory animals   | NOD-SCID or nude (Ncr-Foxn1nu) mice, P35, female. |
|-----------------------|---------------------------------------------------|
| Wild animals          | The study did not involve wild animals             |
| Field-collected samples | The study did not include field-collected samples |

### Human research participants

**Policy information about studies involving human research participants**

| Population characteristics | Paediatric high grade glioma and DIPG patients, defined as ages 0.1-30 years, arising in the cerebral hemispheres, pons or other midline structures. |
|-----------------------------|-----------------------------------------------------------------------------------------------------------------------------------|
| Recruitment                 | Retrospective series across multiple centres for which material was available for genomic analyses. Prospectively from biopsy, surgery or autopsy from which cell culture could be established. |