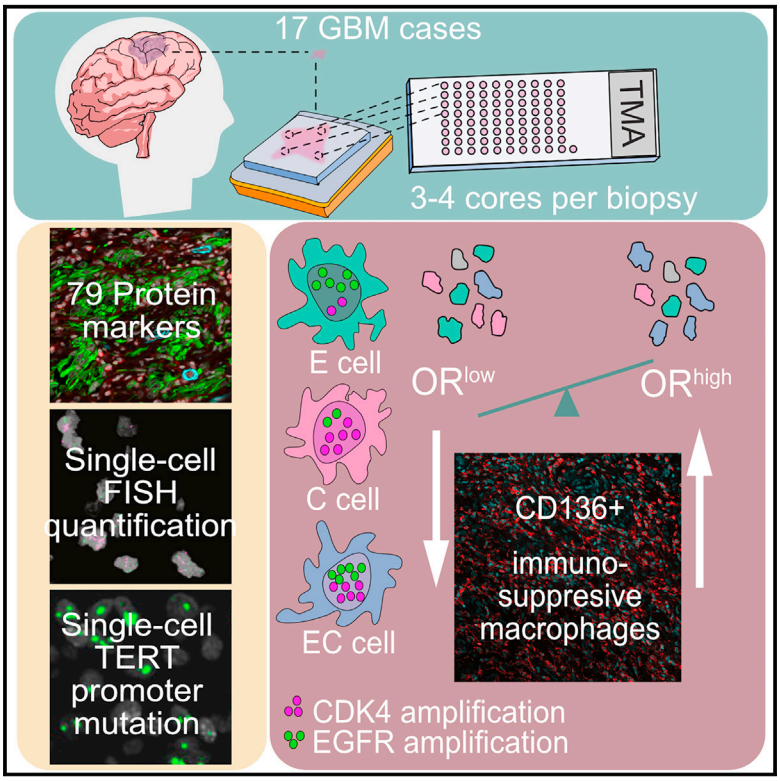
Cell Reports

Article

Single-cell heterogeneity of *EGFR* and *CDK4* co-amplification is linked to immune infiltration in glioblastoma

Graphical abstract



Highlights

- Genetic mosaicism is one of the hallmarks of GBM intratumor heterogeneity
- Multiplexed FISH and spatial profiling resolve genetic diversity in native TME
- Odds ratio for *EGFR* and *CDK4* co-amplification is linked to immune infiltration

Authors

Kacper A. Walentynowicz, Dalit Engelhardt, Simona Cristea, ..., Thomas O. McDonald, Franziska Michor, Michalina Janiszewska

Correspondence

michor@jimmy.harvard.edu (F.M.), mjaniszewska@ufl.edu (M.J.)

In brief

Glioblastoma cellular diversity in the context of tumor tissue microenvironments remains poorly understood. Walentynowicz et al. use the relative frequency of cells harboring co-amplification of *EGFR* and *CDK4* to classify tumors and show that the presence of genetically distinct subpopulations is associated with differences in immunosuppressive macrophage infiltration in glioblastoma tissue.

Walentynowicz et al., 2023, Cell Reports *42*, 112235
March 28, 2023
https://doi.org/10.1016/j.celrep.2023.112235

CellPress

Cell Reports

CellPress
OPEN ACCESS

Article

Single-cell heterogeneity of *EGFR* and *CDK4* co-amplification is linked to immune infiltration in glioblastoma

Kacper A. Walentynowicz,[1,2,12,14] Dalit Engelhardt,[3,4,5,6,14] Simona Cristea,[4,5,6,7,14] Shreya Yadav,[1,2,12] Ugoma Onubogu,[1,2,8] Roberto Salatino,[1,2,8] Melanie Maerken,[1,2,13] Cristina Vincentelli,[9] Aashna Jhaveri,[4] Jacob Geisberg,[4] Thomas O. McDonald,[3,4,5,6] Franziska Michor,[3,4,5,6,10,11,*] and Michalina Janiszewska[1,2,8,15,*]

[1]Department of Molecular Medicine, The Herbert Wertheim UF Scripps Institute for Biomedical Innovation and Technologies, Jupiter, FL, USA
[2]Department of Molecular Medicine, Scripps Research, Jupiter, FL, USA
[3]Center for Cancer Evolution, Dana-Farber Cancer Institute, Boston, MA, USA
[4]Department of Data Science, Dana-Farber Cancer Institute, Boston, MA, USA
[5]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA
[6]Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA, USA
[7]Department of Medical Oncology, Harvard Medical School, Boston, MA, USA
[8]The Skaggs Graduate School of Chemical and Biological Sciences, The Scripps Research Institute, La Jolla, CA, USA
[9]Department of Pathology, Mount Sinai Medical Center, Miami Beach, FL, USA
[10]The Broad Institute of MIT and Harvard, Cambridge, MA, USA
[11]The Ludwig Center at Harvard, Boston, MA, USA
[12]Present address: Memorial Sloan Kettering Cancer Center, New York, NY, USA
[13]Present address: Synthego Corporation, Redwood City, CA, USA
[14]These authors contributed equally
[15]Lead contact
*Correspondence: michor@jimmy.harvard.edu (F.M.), mjaniszewska@ufl.edu (M.J.)
https://doi.org/10.1016/j.celrep.2023.112235

SUMMARY

Glioblastoma (GBM) is the most aggressive brain tumor, with a median survival of ~15 months. Targeted approaches have not been successful in this tumor type due to the large extent of intratumor heterogeneity. Mosaic amplification of oncogenes suggests that multiple genetically distinct clones are present in each tumor. To uncover the relationships between genetically diverse subpopulations of GBM cells and their native tumor microenvironment, we employ highly multiplexed spatial protein profiling coupled with single-cell spatial mapping of fluorescence *in situ* hybridization (FISH) for *EGFR*, *CDK4*, and *PDGFRA*. Single-cell FISH analysis of a total of 35,843 single nuclei reveals that tumors in which amplifications of *EGFR* and *CDK4* more frequently co-occur in the same cell exhibit higher infiltration of $CD163^{+}$ immunosuppressive macrophages. Our results suggest that high-throughput assessment of genomic alterations at the single-cell level could provide a measure for predicting the immune state of GBM.

INTRODUCTION

Glioblastoma (GBM) is the most heterogeneous and aggressive brain malignancy, with an average survival of 15–18 months post-diagnosis.[1] Even through GBM's frequency of 3 cases per 100,000 people in the US is moderate, this disease causes high morbidity and mortality, as only 6.8% patients survive beyond 5 years post-diagnosis.[1] Despite a growing understanding of the disease biology, therapies targeting molecular features in GBM have been failing in clinical trials.[2]

Genetically, GBM is characterized by a complex mutational landscape with a large degree of inter- and intratumor heterogeneity.[3–5] Gains in chromosome 7 and loss of chromosome 10 are common events predicted to arise early in tumor evolution.[6] Frequent amplifications of several receptor tyrosine kinases (RTKs), including EGFR, PDGFRA, and MET, have been explored as potential therapeutic targets in GBM.[1] However, despite successful trials in other malignancies, RTK inhibitors have not provided significant benefit to patients with GBM.[1,7,8] In addition to the adaptability of cellular signaling in response to RTK inhibition, and the challenges associated with drug penetration through the blood-brain barrier, genetic intratumor heterogeneity further contributes to GBM resistance.[9] In particular, RTK-encoding gene amplifications, particularly in *PDGFRA*, *EGFR*, and *MET,* frequently co-occur in the same GBM sample, yet not in the same cells.[10,11] Thus, genetic mosaicism of GBM is one of the key obstacles for effective treatment of these tumors.

Inter- and intratumor heterogeneity of gene amplifications is also linked to the diversity of phenotypes identified in GBM. Initial transcriptional profiling of bulk tumor tissue revealed that GBMs

can be classified into three subtypes: classical-like, proneural, and mesenchymal.[12,13] Each of these transcriptional subtypes is associated with increased frequency of specific genetic alterations: *EGFR* amplification and loss of *CDKN2A/B* are typical for classical-like GBM, *CDK4* and *PDGFRA* amplifications are common for proneural GBM, and *NF1* loss is associated with a mesenchymal transcriptional program.[13] Despite this discovery, the restriction to the proneural tumor type in clinical trials for PDGFRA inhibition did not increase the treatment success rate.[1] Advances in single-cell transcriptomics revealed that each GBM tumor, regardless of subtype reported by bulk profiling, is in fact a mixture of cells belonging to different subtypes.[14–17] One of the most recent classifications of cells based on single-cell expression profiling identified four major cellular states, recapitulating distinct neural cell types.[15] This classification also points to gene amplifications as potent drivers of intratumor heterogeneity, as each of the cellular states is associated with a distinct genetic alteration: oligodendrocytic precursor cell-like (OPC-like) with *PDGFRA* amplification; neural progenitor cell-like (NPC-like) with *CDK4* amplification; astrocytic cell-like (AC-like) with *EGFR* amplification; and mesenchymal-like (MES-like) with NF1 loss.[15] While transitions between these cell types can occur, each genetic driver favors a particular transcriptional state of the cell.[15] Thus, assessment of copy-number alterations (CNAs) in *EGFR*, *CDK4*, and *PDGFRA* at the single-cell level *in situ* may provide information about cellular diversity within distinct tumor microenvironments.

In situ heterogeneity of GBM, observed as extensive variability of histological features, is characteristic for this tumor type and forms the root of the historical name "glioblastoma multiforme." The disordered tissue structure as well as non-uniform contrast enhancement in magnetic resonance imaging (MRI) point to a high degree of macroscopic diversity within these tumors.[18] Characteristic pathological features of GBM include areas of pseudopalisading, organized alignment of viable cells surrounding the necrotic regions, and microvascular proliferation.[19] These different tumor microenvironmental types are often intermixed in the tumor tissue. Interestingly, both the perinecrotic areas and perivascular regions were shown to support GBM stem-like cells, rare subpopulations with increased epigenetic plasticity and potentially higher resistance to standard treatment.[20–24] Thus, to survive, GBM cells may need to adapt to drastically different microenvironments. Yet, this adaptation is not necessarily cell autonomous; instead, the maintenance of intratumor heterogeneity in GBM may be dependent on interactions between genetically diverse GBM cell subpopulations and may be driven by differences in and interactions with specific microenvironmental components.

Since the four major subtypes of GBM cells, while remaining plastic, are largely determined by underlying genetic aberrations,[15] we aimed to study how genetically constricted populations interact with each other in the context of the tumor microenvironment of intact clinical tissue specimens.

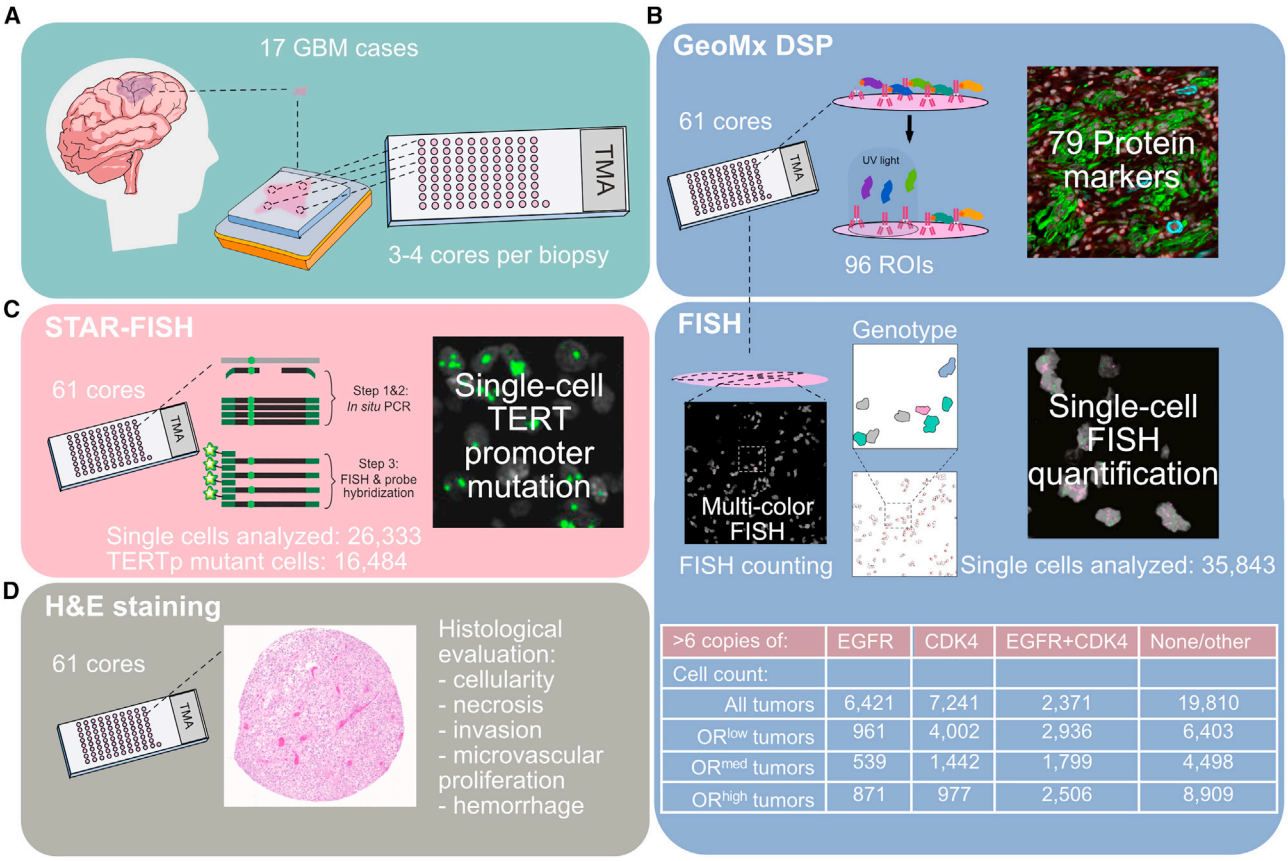
RESULTS

To investigate the relationship between the spatial localization of cells with different genotypes and their microenvironmental preferences, we constructed a tissue microarray (TMA) from 17 formalin-fixed paraffin-embedded (FFPE) IDH wild-type GBM cases (Table S1; all tumor samples were collected prior to treatment). From each tumor block, 3–4 cores were randomly selected (61 in total) to investigate the patterns of local intratumor heterogeneity within each tumor biopsy (Figure 1A). Since these samples were of variable age and embedded in paraffin blocks, they could not be used for transcriptomic studies; however, formalin-fixed paraffin embedding and prolonged storage in ambient temperature does not impede immunohistochemistry or DNA-based studies. Thus, we used our multiregion TMA of intact tissue samples from each tumor for single-cell assessment of selected genomic features by fluorescence *in situ* hybridization (FISH) and specific-to-allele PCR-FISH (STAR-FISH) as well as protein expression using NanoString's GeoMx digital spatial profiling (DSP) platform (Figure 1).

GeoMx DSP data reveal local heterogeneity in tumor microenvironmental marker expression

To characterize the extent of variation in the local GBM tumor microenvironment within distinct regions of the same tumor biopsy, we employed a spatial profiling platform based on highly multiplexed immunostaining with oligonucleotide-tagged antibodies (Figure 2A). GeoMx DSP allows quantification of oligo tags released by UV-illuminated regions of interest (ROIs), which can be selected based on immunofluorescent staining for a few markers. DSP enables simultaneous antibody staining for tens of markers and control antibodies, providing more robust quantification compared with fluorescence-based methods. In our assay, multicolor immunofluorescent staining with markers of glial cells (GFAP), immune cells (CD45), and vasculature (α-SMA) was used to select ROIs within each TMA core (Figure S1A). Despite the small size of each core (1 mm in diameter), many of the cores display varied morphology and expression levels for GFAP, CD45, and α-SMA within the core (Figure 2B). We selected 96 ROIs out of the 61 cores present on the TMA (see STAR Methods), with multiple (2–3) ROIs in cores with heterogeneous immunofluorescent staining (Figure S1A). These ROIs were then subjected to GeoMx DSP for 79 protein markers associated with GBM biology, neurobiology, and immunology (Figure 2C; Table S2). We then performed a multilevel analysis of correlation in expression patterns across tumors and a quantification of intra- vs. intertumor expression heterogeneity based on ROI- and core-specific expression patterns.

To identify trends in co-expression, we assessed overall correlations between protein expression across tumors via hierarchical clustering (Figure 2C, clusters of proteins with similar function are indicated on the figure). We found that the expression levels of the immune markers CD68, HLA-DR, CD11b, CD45, IBA1, CD163, and CD14 were correlated across all tumors (mean Spearman correlation = 0.76, range: 0.57–0.92; Figure 2C, the second cluster from bottom right). A similar trend was observed for the neuron and glial cell-expressed TAU protein and its phosphorylated versions, except for phosphor-TAU S404 (Spearman correlation > 0.69; Figure 2C, the third cluster from top left). Negative correlation was observed between clusters of immune (including CD68, HLA-DR, and CD11b) and neuronal markers (Park5, SNCA-filament, SYP,



>6 copies of:	EGFR	CDK4	EGFR+CDK4	None/other
Cell count:				
All tumors	6,421	7,241	2,371	19,810
OR^{low} tumors	961	4,002	2,936	6,403
OR^{med} tumors	539	1,442	1,799	4,498
OR^{high} tumors	871	977	2,506	8,909

Figure 1. Summary of study design and acquired data

(A) Schematic of tumor microarray (TMA) construction.

(B) TMA slide 1 was first used for GeoMx digital spatial profiling and next for multicolor FISH. Quantification of individual FISH signals was performed to classify cells into distinct genotypes based on *EGFR* and *CDK4* amplification. Table contains overall cell count for each genotype. OR^{low}, tumors with low odds ratio for *CDK4*/*EGFR* co-amplification at single-cell level; OR^{high} tumors, high co-amplification odds ratio.

(C) TMA slide 2 was used for STAR-FISH for *TERT* promoter hotspot mutation detection.

(D) TMA slide 3 was used for H&E-based histological evaluation.

and MBP, among others), with a Spearman coefficient of −0.48, even though this comparison did not reach statistical significance ($p = 0.052$; see STAR Methods). Strong correlations were also observed between the expression of Lef1 and Rspo2, two proteins of the Wnt signaling pathway, and Satb2, a protein of the TCF/LEF pathway (Spearman correlation coefficients ranging from 0.74 to 0.93). These proteins have been implicated in stemness and promotion of a cancer stem cell phenotype.[25–27] Thus, our DSP delivers expected protein co-expression patterns.

The expression levels of the 79 markers were highly variable among cores derived from the same tumor (Figure S1B). Since the TMA cores represent spatially distant regions of the same tumor fragment with possibly different compositions, we assessed the variability of protein expression between ROIs from distinct tumors relative to the variability within tumors using the Kruskal-Wallis H test. Overall, we found that the variability of our 79 protein marker expression panel between tumors exceeded the variability within tumors (Figure 2D). Particularly high inter- relative to intratumor variability was observed for several immune-related proteins (CD45, HLA-DR, IBA1, P2RX7, YKL-40), as well as for EGFR and p53, which are both frequently overexpressed and mutated in human GBM[4,28] and can drive tumorigenesis in animal models of glioma.[29,30] This finding may indicate that both immune state and major tumor drivers are less variable between biopsies of the same tumor. The neuronal markers NEFL, MBP, and Park5 also exhibited higher relative intertumor variance. On the other hand, vascular cell adhesion molecule (VCAM) and CD31, both associated with endothelial cell function and tumor vascularization,[31] exhibited comparatively high expression variance between regions within the same tumor sample and among samples, suggesting that, within GBM biopsies, there is high variability in blood vessel distribution and that even a single biopsy is likely to contain differentially vascularized niches (Figures S2A and S2B).

This result prompted us to assess the histopathological features of each TMA core to identify which protein markers may be specific for different tissue compositions (Figure S2). Tissue composition assessment based on H&E staining confirmed the expected variation in hemorrhage and necrosis areas within

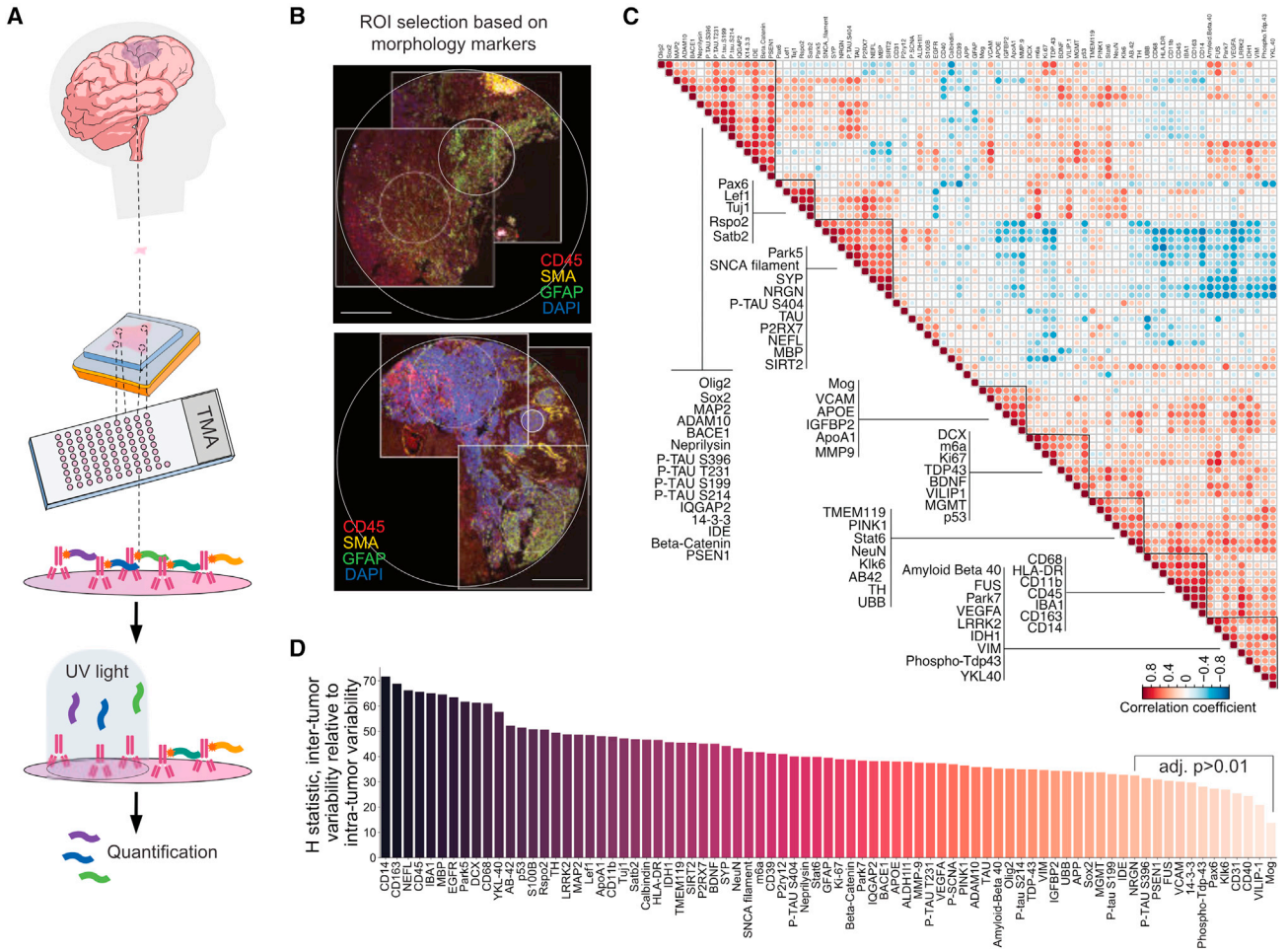


Figure 2. Spatial profiling reveals regional heterogeneity in protein expression between distinct areas of the same GBM biopsy

(A) Schematic of digital spatial profiling (DSP) for protein marker expression. From each FFPE block, 3–4 cores were punched to construct a multiregion TMA. In DSP assay, staining with 79 protein marker-specific antibodies labeled with photocleavable oligo tags were used. UV light exposure of a region of interest (ROI) releases the tags, which are then quantified for each ROI separately.

(B) Example of two cores with low (left) and high (right) heterogeneous staining, used to select ROI for DSP. Large circle: full core size (1 mm diameter), small circles: selected ROIs. Scale bar, 20 μm.

(C) Spearman correlations between all profiled markers across tumors. Groups of markers with the strongest correlations are marked.

(D) Relative measure of inter- to intratumor variability in the expression of each protein, as indicated by the H Kruskal-Wallis statistic (see STAR Methods).

cores from the same tumor (Figure S2C). However, classification of each core according to the most dominant histopathological feature did not reveal visible protein expression correlation patterns (Figure S2D), possibly due to cores containing mixed features. Histopathological features, especially the immune infiltration, can be affected by anti-inflammatory treatment with steroids, which are often given to patients with GBM before surgical resection of the tumor to ease edema symptoms.[1,32–34] In our cohort, only 4 cases received steroids, and we did not observe any specific patterns of protein expression or histopathological features in these tumors (Figure S2D).

Characterization of single-cell genetic heterogeneity in GBM multicore TMA

The analysis of the spatial localization of cells harboring genetic drivers associated with distinct transcriptional states in GBM, namely amplifications of *EGFR*, *PDGFRA*, and *CDK4*,[15] could be revealing of microenvironmental preferences and interactions between these cellular subpopulations. To address this question, we performed multiplexed FISH for *EGFR*, *PDGFRA*, and *CDK4* on our multiregion TMA (Figures 3A and S3A). Using an automated single-cell FISH signal counting platform, we quantified the number of FISH signals in each individual nucleus imaged (n = 35,843 nuclei, on average 2,335 nuclei per tumor, 216 nuclei per image, 166 images). We found that *PDGFRA* amplification (≥6 FISH signals) was relatively rare, detected in less than 2% of all cells analyzed, and we therefore focused on the amplification patterns of *EGFR* and *CDK4*, independently of *PDGFRA* amplification status. Each cell was assigned to one of the following genotypes: E (*EGFR* amplified and not *CDK4* amplified), C (*CDK4* amplified and not *EGFR* amplified), EC (*EGFR* and *CDK4* amplified), and N/O (no amplification of

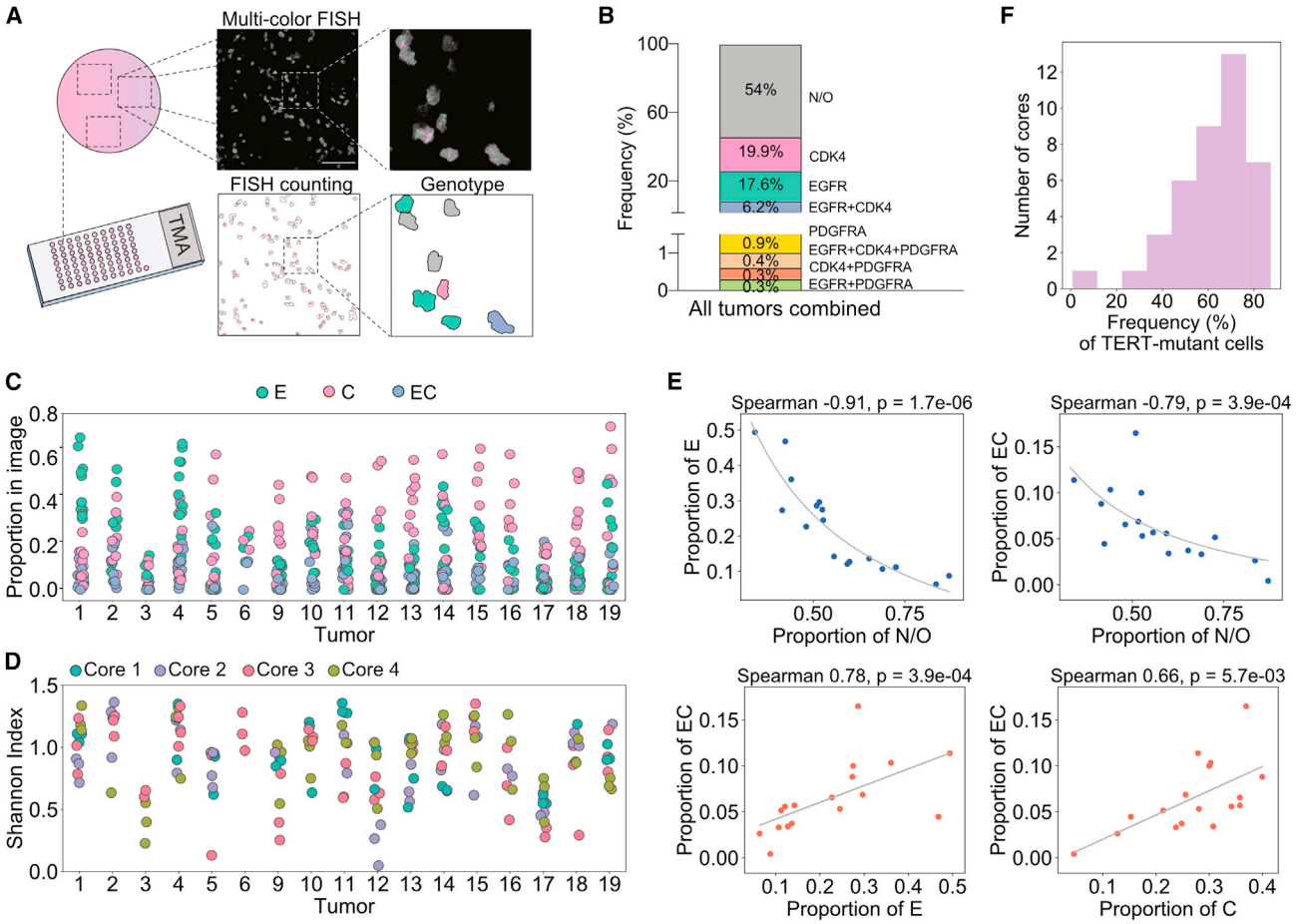


Figure 3. Single-cell characterization of *EGFR*, *PDGFRA*, and *CDK4* amplifications and *TERT* promoter mutations in a multiregion GBM TMA

(A) Summary of the FISH quantification workflow. Each imaged area is segmented, based on the nuclear outlines, for counting of each FISH signal in individual nuclei. Based on the presence of gene amplifications, each nucleus is classified as a distinct genotype. Scale bar, 50 μm.

(B) Overall frequency of cells with FISH-based derived genotypes.

(C) Proportion of cells with the E, C, and EC genotypes in individual images in each tumor. E, cells with amplified *EGFR* and no *CDK4* amplification; C, cells with amplified *CDK4* and no *EGFR* amplification; EC, cells with amplification of both *EGFR* and *CDK4*. Each data point represents a genotype frequency in an individual image.

(D) Heterogeneity of cells (measured by the Shannon index) across the E, C, EC, and N/O (not amplified in either *CDK4* or *EGFR*) genotypes within each tumor. Each data point represents an image. Colors represent different cores of the same tumor.

(E) Correlations between proportions of cells with distinct genotypes, with each point representing the ratio of the cells with a particular genotype to the total number of cells within a tumor. Spearman correlation coefficients and p values adjusted for multiple comparisons are shown along with best-fit curves. Scales allowing for best visualization of data points (and not showing the full 0–1 proportion ranges) were used in each plot.

(F) Within-core frequency of cells harboring *TERT* promoter mutation.

either *EGFR* or *CDK4*). We observed significant variation in the overall frequency of cells of each genotype within each TMA tumor (Figures 3B–3D; Table S3). Although several tumors had relatively high E cell frequency, C and EC cells were found in most of the analyzed samples (Figure 3C). A calculation of the Shannon index of diversity[35] of genotype prevalence within each image (Figure 3D) showed variation between distinct cores taken from the same biopsy, further suggesting that the extent of heterogeneity differs within relatively small tumor regions. We did not identify any relationships between a tumor's anatomical location and its genetic diversity (Figure S4), a finding that may be due to the small size of our cohort, enriched in frontal and temporal lobe tumors.

About half of all cells quantified in our samples were identified as N/O. This cell population could include cancer cells with distinct genotypes, without high copy-number amplification of the *EGFR* and *CDK4* genes, but with other genetic alterations. It could also include cells of the tumor microenvironment that have a normal (non-amplified) genotype. Correlation of N/O cell frequency with DSP-based protein expression did not reveal a distinct identity of this population (Figure S3B). However, we found a negative correlation between the frequency of N/O cells and expression of Olig2 (Spearman correlation = −0.75, adjusted p value = 0.044). Our protein expression profiling found Olig2 expression to be strongly correlated with Sox2 expression (Spearman coefficient = 0.81, adjusted p value = 0.007),

although within the N/O population, the negative correlation with Sox2 expression did not reach statistical significance. Both Olig2 and Sox2 are often expressed at high levels by GBM tumor cells,[12,13] and thus a low abundance of these proteins in N/O cells could indicate that a large proportion of these cells is of non-tumor origin.

Since cells with distinct genotypes based on the three selected amplifications were found in almost all tumors analyzed, we investigated whether there were any relationships between the proportions of these cell types in a tumor (Figures 3E and S3C). The prevalence of N/O cells in a tumor was negatively correlated with E cell frequency (Spearman correlation = −0.91, adjusted p value = 1.7×10^{-6}) and EC cell frequency (Spearman correlation = −0.79, adjusted p value = 3.9×10^{-4}), both of which were positively correlated with each other (Spearman correlation = 0.78, adjusted p value = 5.7×10^{-3}). EC cell frequency was also positively correlated, albeit more weakly, with C cell frequency (Spearman coefficient = 0.66, adjusted p value = 5.7×10^{-3}). No strong correlation was found between C and E cell frequency, and a weak correlation was identified between C and N/O cell frequency (Spearman coefficient = −0.52, adjusted p value = 0.04; Figure S3C).

Next, we sought to establish whether the tumor genotypes based on bulk tissue sequencing could be linked to diversity of single-cell-based, copy-number-derived genotypes. Targeted sequencing of 50 genes using the OncoPanel (see STAR Methods) showed only a few mutations present in each tumor in our cohort, and no clear distinction between tumors could be made based on these genetic alterations assessed in bulk (Figure S3D; Table S4).

hTERT promoter hotspot mutations have been reported as one of the drivers of GBM,[4,36] yet they were not included in the targeted sequencing panel we used. To assess the frequency and spatial localization of cells harboring the hotspot mutation C228T in the *hTERT* promoter region, we performed *in situ* single-cell mutation detection using the STAR-FISH assay[37] on our multiregion GBM TMA (Figure S5). We found that the majority of the cores in the TMA contain over 50% TERT mutant cells, hetero- and/or homozygous (Figure 3F). There was significant variation between distinct cores taken from different regions of the same tumor (Figure S5D). We did not observe any correlations between the ratio of cells with *hTERT* promoter mutation and the frequency of N/O cells, i.e., cells lacking *EGFR*, *CDK4*, and *PDGFRA* amplifications, most likely due to the heterogeneous nature of the N/O population (Figure S5E).

Together, these results show that single-cell-based FISH quantification identifies a high prevalence of *hTERT* mutations and of cells harboring *CDK4*, *EGFR*, and dual amplifications. These genetic alterations co-exist in the majority of the GBM samples tested and exhibit a large extent of heterogeneity.

Single-cell genotypes stratify tumors by co-occurrence of EGFR and CDK4 in the same cell vs. distinct cell populations

Since amplifications of *EGFR*, *PDGFRA*, and *CDK4* were previously shown to correlate with the presence of transcriptionally distinct cell states in GBM,[15] cells harboring a combination of these genetic events could gain dual properties or have new distinct phenotypes. Our dataset of single-cell genotypes allows for an investigation of differences between tumors in which these alterations tend to co-occur in the same cell and tumors in which they are more likely to be found in distinct cell populations. We therefore calculated the odds ratio of detecting co-amplified cells within a tumor for each of our 17 GBM cases. Since *PDGFRA* amplification was rare in our dataset, we again focused on *EGFR* and *CDK4* amplifications. The odds ratio (OR), as defined here, represents the likelihood that a cell harboring an *EGFR* amplification also harbors a *CDK4* amplification, relative to the likelihood of a *CDK4* amplification unaccompanied by an *EGFR* amplification. At OR = 1, the presence of an amplification of one of these genes has no effect on the probability of the other gene being amplified. An OR above (below) unity therefore implies an increased (decreased) tendency for *EGFR* and *CDK4* amplifications to co-occur in the same cell over what would be expected if the presence of one amplification had no effect on the presence of the other. The OR calculation (see STAR Methods) accounts for all relative frequencies of the cells with distinct genotypes (C, E, CE, N/O) within a tumor. Using the OR values, we separated the tumors into 3 classes based on OR tertiles (n = 6, n = 5, and n = 6): OR^{low}, characterized by a decreased tendency for *EGFR* and *CDK4* same-cell co-amplification (OR is significantly below unity); OR^{high}, characterized by an increased tendency for same-cell co-amplification (OR is significantly above unity); and tumors where there is no strong effect either way (OR near unity) (Figures 4A and 4B; Table S5). The frequencies of E cells (*EGFR* amplification alone) and N/O cells (lacking *EGFR* and *CDK4* amplifications) were significantly different between OR^{high} and OR^{low} groups, while C and EC cells were similarly distributed across all tumors (Figure 4C). Thus, the low OR may be driven by an increased frequency of cells harboring only the *EGFR* amplification. While the median frequency of cells with an *hTERT* promoter mutation was similar across the OR groups, OR^{low} tumors had somewhat higher variability in *hTERT* mutation frequency, albeit not significantly, which could arise if the mutation was subclonal in these tumors. Interestingly, the average copy number of *EGFR* in E cells was significantly higher in OR^{low} tumors (Figure 4D; mean 9.7 copies in OR^{low} vs. 7.8 copies in OR^{high}, Mann-Whitney test p value = 0.015) and also exhibited higher variability in OR^{low} tumors, as indicated by the Shannon diversity index (Figure 4F). In contrast, *CDK4* copy-number diversity remained similar among tumors with different OR statuses (Figures 4E and 4F). This suggests that the nature of *EGFR* amplification in E cells in OR^{low} tumors may be qualitatively different from that observed in E cells in OR^{high} tumors. A possible explanation is that *EGFR* amplification could be generated by extrachromosomal DNA fragments in OR^{low} tumors. To test this hypothesis, we performed whole-genome sequencing of DNA extracted from FFPE slides from three OR^{low} and three OR^{high} tumors. Standard CNA analysis confirmed high levels of *EGFR* amplification in all three OR^{low} tumors, consistent with our FISH-based findings (Figure S6A). In the OR^{high} group, we confirmed *CDK4* amplification in one tumor, where a large clonal fraction of C and EC cells was expected. However, the co-amplification of *EGFR* and *CDK4* in OR^{high} tumors was not detected in bulk sequencing. This discrepancy is not unexpected given the subclonal nature of

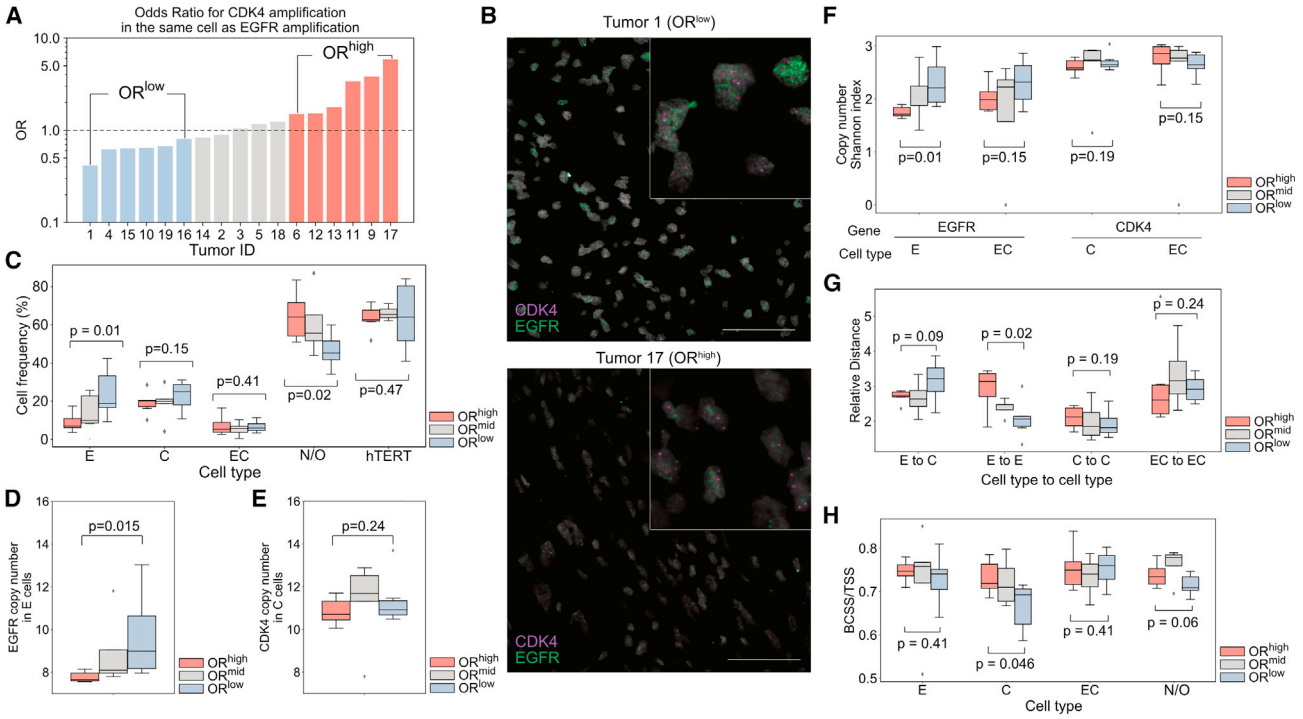


Figure 4. Single-cell genotypic classification of tumors based on relative proportions of cells with *EGFR* and *EGFR+CDK4* amplifications

(A) Odds ratios (ORs) for the presence of *CDK4* amplification in the same cell as *EGFR* amplification, relative to cells with *CDK4* amplification alone. OR tertiles are shown in different colors.

(B) Representative images of *EGFR* and *CDK4* FISH in tumors with low (tumor 1) and high (tumor 17) ORs. Gray: DAPI, green: *EGFR* FISH, magenta: *CDK4* FISH. Scale bar, 50 μm.

(C) Frequencies of E, C, EC, and N/O cells as well as cells harboring *TERT* promoter mutations in low- and high-OR tumors.

(D) Per-tumor mean *EGFR* copy number in E cells across the OR groups.

(E) Per-tumor mean *CDK4* copy number in C cells across the OR groups.

(F) Diversity (per-tumor Shannon index) of *EGFR* and *CDK4* CNA in single- and dual-amplified cells across the OR groups.

(G) Within-image distances between cells of different genotypes relative to distances among N/O cells across OR groups.

(H) Clustering of cells with distinct genotypes. Between cluster sum of squares by total sum of squares (BCSS/TSS; see STAR Methods) higher ratio indicates tighter clustering. The box-and-whisker plots in all bar graphs (C–H) show the mean (midline) and 25th–75th (box) and 5th–95th (whiskers) percentiles. Mann-Whitney tests p values are shown. Colors of tumor groupings are indicated in (A). Genotypes: E, cells with amplified *EGFR* and non-amplified *CDK4*; C, cells with amplified *CDK4* and non-amplified *EGFR*; EC, cells with amplification of both *EGFR* and *CDK4*; N/O, cells not amplified in either *EGFR* or *CDK4*. Each point represents a single tumor, with weighted averages (by cell number) for image-wise quantities (distances and clustering).

the co-amplifications identified by our single-cell-based FISH counting. Next, to test for the presence of structural signatures suggestive of a potential extrachromosomal nature of *EGFR* amplifications, we reconstructed the fine structure within the focal amplification of *EGFR* using the AmpliconArchitect.[38] Indeed, in the three OR^{low} tumors tested, we observed a complex structure of genomic rearrangements involving the *EGFR* amplicon, which stands in contrast to the simpler amplification patterns found in the three OR^{high} tumors (Figure S6B). The complex breakpoints identified in our whole-genome sequencing data resemble findings of ApliconArchitect reconstruction of extrachromosomal DNA (ecDNA) confirmed by BioNano optical mapping.[39] Thus, both our FISH and whole-genome sequencing data point to an extrachromosomal nature of EGFR amplification in OR^{low}, but not in OR^{high}, tumors.

Our *in situ* single-cell genotyping dataset contains spatial coordinates for each nucleus recorded in an image. This dataset thus enabled us to compute the distance between spatial locations of cells with distinct genotypes in OR^{low} and OR^{high} tumors. Compared with OR^{high} tumors, E cells in OR^{low} tumors are closer to each other (Mann-Whitney test p value = 0.02; Figure 4G), while no significant difference in distances among C cells was found between OR^{low} and OR^{high} tumors (distances shown are relative to N/O-N/O cell distances in order to normalize for variability in tissue density). Since the difference in abundance of E cells in OR^{low} and OR^{high} tumors is higher than that of C cells (Figure 4G), the proximity of E cells may be directly related to frequency of these cells. The ratio among between-cluster sum of squares and total sum of squares, which informs about the clustering of individual cell types, showed that in OR^{high} tumors, C cells have a tendency toward tighter clustering within the tissue (Mann-Whitney test p value = 0.046; Figure 4H). This and the lower relative distance between C-to-C cells compared with other cell types (Figure 4G) suggest that cells harboring only the *CDK4* amplification could potentially be less migratory and stay closer to each other after cell division, creating pockets of C cells.

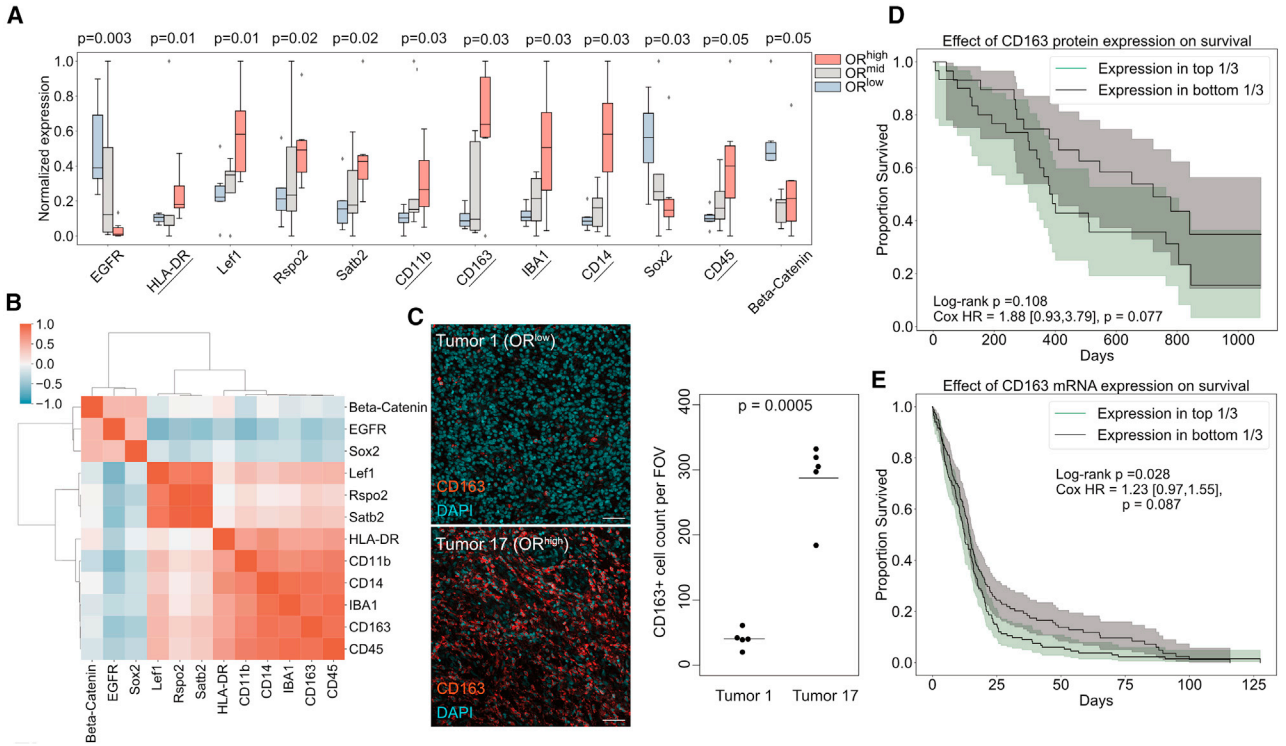


Figure 5. *EGFR* and *CDK4* co-amplification OR-based tumor groups display distinct protein expression patterns associated with immune infiltration and survival outcomes

(A) Differential protein expression between OR-based tumor groups. The expression value of a protein in a tumor is a weighted mean of expression levels in individual ROIs in that tumor. Each bar represents expression values across all tumors in a group, the mean (midline), and 25th–75th (box) and 5th–95th (whiskers) percentiles are shown. Only proteins exhibiting an adjusted Mann-Whitney test p value <0.05 are shown. Underlined: immune markers.

(B) Tumor-wise correlations between the differentially expressed proteins in (A).

(C) Representative images of CD163 staining in the tumors with the highest and lowest ORs (left) and CD163^{+} cell quantification (right). FOV, field of view. Scale bar: 50 μm.

(D) Survival analysis for patients stratified into top and bottom tertiles (n = 31 and n = 30, respectively; CPTAC[40] dataset) based on CD163 protein level. Log rank test p values and Cox proportional hazard ratio (HR) values (age and sex adjusted; see STAR Methods) are shown with a 95% confidence interval in brackets. Shaded areas: confidence intervals.

(E) Survival analysis for patients stratified into top and bottom tertiles (n = 170 for each group, TCGA Firehose Legacy dataset) based on CD163 mRNA expression level. Log rank test p values and Cox proportional HR values (age and sex adjusted; see STAR Methods) are shown with a 95% confidence interval in brackets. Shaded areas: confidence intervals.

In summary, our cytogenetic and spatial analysis showed that the single-cell-level quantification of *EGFR* and *CDK4* amplifications can be used to stratify tumors into distinct classes based on the relative frequency of cells with co-amplification of the two oncogenes: OR^{low} tumors, characterized by higher relative presence of *EGFR*-only amplified cells, and OR^{high} tumors, with more balanced frequency of cells with single and both amplifications.

Single-cell EGFR-CDK4 amplification co-occurrence is associated with an immunosuppressive tumor microenvironment

Since the OR^{low} and OR^{high} tumor groups exhibit distinct cellular composition and distributions of *EGFR* and *CDK4* gains, we sought to identify whether any proteins are differentially expressed between these tumor groups. As expected, the levels of the EGFR protein were significantly different between OR^{low} and OR^{high} tumors (Mann-Whitney test p value = 0.03; Figure 5A), reflecting the differences in the relative numbers of cells harboring *EGFR* gene amplification as well as the copy-number differences noted above. More interestingly, we found that OR^{high} tumors were highly enriched in proteins associated with immune cell infiltration and an immunosuppressive microenvironment (CD163, IBA1, CD14, CD45, CD11b, HLA-DR; Figure 5A), all downregulated in OR^{low} tumors. These markers additionally showed a trend toward negative correlations with EGFR protein expression (Figure 5B), although these correlations were not statistically significant. Of note, CDK4 was not included in the GeoMX DSP panel, and therefore no protein expression for this gene was available for our analysis.

Among the immune-related proteins in our expression panel, CD163, a marker of immunosuppressive macrophages, showed the highest mean difference in expression between the OR^{high} and OR^{low} tumor groups, suggesting enrichment of immunosuppressive cells in OR^{high} tumors (Figures 5A and 5B). Immunofluorescent staining of the original sections from OR^{low} and OR^{high} tumors confirmed the striking difference in infiltration of

$CD163^+$ cells (Figure 5C). These results suggest that tumors classified based on the OR of *EGFR* and *CDK4* co-occurrence at the single-cell level have different tumor microenvironments and that a higher relative frequency of co-amplification of these two genes is associated with a more immunosuppressive state. The presence of immunosuppressive macrophages and CD163 expression in tumors has been previously linked with the survival of patients with GBM.[41,42] In line with this observation, using CPTAC[40] and TCGA[4] datasets on CD163 protein and transcript level-based stratifications, respectively, we observed that patients with GBM with higher CD163 expression tend to have somewhat shorter survival times; however, in these datasets, the differences were not significant (Figures 5D and 5E).

To further investigate the extent of *EGFR* and *CDK4* co-amplification in other datasets, we analyzed two large, published human GBM single-cell transcriptomic datasets,[15,16] filtered for IDH wild-type (IDHWT) tumors only. We estimated single-cell-level copy-number gains and losses in genomic regions spanning genes around *EGFR* on chromosome 7 and *CDK4* on chromosome 12 in malignant cells[14] and classified cells accordingly as E, C, or EC cells (Figures S7A and S7B). We found that 12 out of 20 tumors in the Neftel et al.[15] dataset were classified as containing only *EGFR*-amplified cells, while 6 tumors contained cells with inferred co-amplifications of both *EGFR* and *CDK4* (Figure S7B, left panel). Thus, single-cell transcriptomic analysis supports our FISH data-based findings of a high frequency of EGFR amplification in patients with GBM. Since the Neftel et al.[15] tumor samples were depleted of immune cells via CD45 sorting, to analyze the immune landscape in GBM, we turned to the Richards et al.[16] dataset, which contain fewer tumors but has high numbers of immune cells sequenced. Tumor composition in this dataset was more heterogeneous, with several tumors predicted to contain both E and EC cells and three tumors containing small fractions of C cells (Figure S7B, right panel). Interestingly, the tumor G910, with the highest frequency of EC cells, also had the highest frequency of M2-like macrophages (Figures S7C–S7E; Tables S6 and S7). These results further support our finding of OR^{high} tumors having a large extent of infiltration of immunosuppressive macrophages; however, single-cell profiling of larger cohorts containing both tumor and immune cells will be needed to validate this finding.

Analysis of human tumor tissues provides a static snapshot of tumor heterogeneity. Thus, to investigate the possible mechanism by which the OR of *EGFR* and *CDK4* co-amplification might influence the immune microenvironment, we developed a syngeneic murine glioma model based on the GL261 murine cell line. GL261 E, C, and EC populations were generated by overexpressing EGFR and EGFP, CDK4, and mCherry or both oncogene and fluorescent protein sets (Figures S8A and S8B). We aimed to generate an immunocompetent orthotopic model of OR^{low} and OR^{high} tumors by mixing E:C:EC GL261 cells in 73%:19%:8% and 22%:60%:18% ratios, respectively, and injecting the cell mixture into the brain parenchyma of C57BL/6 mice (Figure S8C). However, already at day 21 post-transplantation, the composition of the syngeneic tumors had shifted from the initial clonal mixture, suggesting faster elimination of cells harboring EGFR overexpression (Figure S8D). This finding could be explained by a higher frequency of infiltrating microglia, the resident brain macrophage population, in tumors composed of only EGFR-overexpressing cells compared with tumors generated by WT GL261, C, or EC cells (Figures S8E–S8J). Selection against EGFR-EGFP expression is also supported by a loss of signal from this construct in EC tumors, resulting in the appearance of CDK4-only overexpressing cells (Figure S8D). Additional studies tailored to follow the evolutionary clonal dynamics among E, C, and EC cells with improved clonal retention, possibly by decreasing the immunogenicity of fluorescent protein markers,[43] will be needed to provide more mechanistic insights into the interactions between these genetically distinct clones and the immune microenvironment. Nevertheless, our results suggest a possible divergent interaction between the immune microenvironment and subclones harboring EGFR or CDK4 overexpression.

DISCUSSION

The insight provided by single-cell studies into the diversity of genomic alterations can drive a deeper understanding of the subclonal structure of tumors not afforded by bulk tumor sequencing-based approaches. Key to this understanding is the interaction of a heterogeneous tumor microenvironment with subclonal-level changes.

Our main finding is the identification of the OR of *EGFR-CDK4* co-amplification as an indicator of distinct alteration and expression trends. This quantity is a subclonal-level property that can only be obtained from single-cell data and is hidden by bulk tumor sequencing analyses. High intratumor heterogeneity may especially hamper the ability of bulk sequencing studies to identify prognostically relevant patterns of amplification occurrence and co-occurrence. This fact is exemplified in our study, as *EGFR* and *CDK4* amplifications co-occur in 6.6% of nuclei analyzed, yet in 8 out of the 17 analyzed tumors, the frequency of these dual-amplified cells was more than three times higher than that in at least one imaged region.

Based on TCGA bulk sequencing of GBM genomes, the frequency of co-amplification of *EGFR* and *CDK4* is 8.9% (22/248 IDHWT GBM) and is thus lower than our FISH-based call. Single-cell RNA sequencing (RNA-seq)-based inference confirmed that these subclonal genomic alterations do occur more frequently than expected from bulk tumor data (5/20 and 4/7 tumors in the two cohorts we analyzed[15,16]). It is important to note that these single-cell CNA estimates are based on a sliding window average of the expression values of 100 neighboring genes, centered at the gene of interest.[14,44] Thus, focal amplifications are difficult to infer from single-cell gene expression data alone, which could explain the overall lower diversity of genotypes identified by inferCNV[14,44] compared with what we find experimentally by single-cell FISH quantification. In the future, target-focused DNA-based methods such as MissionBio's Tapestri platform, currently only applicable to fresh frozen tissues, could enable genome-wide genotyping of subclonal co-amplifications of individual genes at the single-cell level.

In our dataset, *EGFR*-amplified cells in tumors with low *EGFR-CDK4* co-amplification OR had a higher average copy number and increased copy-number diversity of *EGFR*, as well

as higher EGFR protein expression. *EGFR* amplification in ecDNA may explain these observations. ecDNA, circularized DNA fragments found in GBM and other tumors, contain an oncogene together with an enhancer element, allowing for increased gene expression.[39,45,46] The circularized form together with the lack of telomeres and centromeres contribute to highly efficient replication of ecDNA, their asymmetrical division between daughter cells in mitosis,[47] and increased heterogeneity, without the fitness penalty associated with large genomic alterations.[48] Moreover, oncogenes within ecDNA can be efficiently overexpressed through protein-tethered ecDNA transcriptional hubs, enabling enhancer input between different ecDNA molecules.[49] In GBM, extrachromosomal amplification of *EGFR* is well documented.[50–52] Here, we also confirmed by whole-genome sequencing analysis the complex structure and heterogeneity of *EGFR*-containing amplicons in selected tumors with a low *EGFR-CDK4* co-amplification OR. Thus, it is possible that our FISH-based analysis identified tumors harboring ecDNA *EGFR* as a distinct class of tumors, associated with low immune marker expression and high levels of EGFR protein.

Our results demonstrate that higher relative frequencies of cells with co-amplification of *EGFR* and *CDK4* (or lower relative frequencies of cells harboring *EGFR* amplification alone) are associated with elevated levels of the CD163 protein, a scavenger receptor expressed in macrophages and monocytes and a classical marker of M2/immunosuppressive polarization.[53] This suggests that the immune cells infiltrating OR^{high} tumors may be predominantly of an immunosuppressive phenotype. The small size of our cohort (n = 5–6 per group) precludes survival analysis; however, high expression of CD163 alone has been previously associated with poor outcome,[41,42,54] and tumor infiltration by the immunosuppressive macrophages is a well-established hallmark of decreased survival for patients with GBM.[19,55]

The link between GBM immune status and underlying genetic alterations in the tumor was explored in several studies. Most notably, inactivating mutations of the *NF1* gene were shown to be associated with increased expression of macrophage-related signatures.[13,56,57] Tumor-infiltrating lymphocytes are depleted and immunosuppressed in tumors representing the classical-like subtype, harboring *EGFR* amplification and *PTEN* deletion.[57,58] These findings are consistent with our observation that tumors with low *EGFR-CDK4* OR, which are also enriched with cells with a high *EGFR* copy number, had lower expression of immune markers. Interestingly, our data show that a higher relative prevalence of cells with dual *CDK4* and *EGFR* amplification was associated with a macrophage-enriched tumor microenvironment. This effect cannot be explained by the presence of *CDK4*, as this amplification alone has been linked to lower macrophage infiltration and fewer $CD4^+$ T cells.[56] Our findings therefore imply that a subpopulation of GBM cells harboring a low level of *EGFR* amplification that co-occurs with amplification of *CDK4* is likely to elicit different immunological effects than subpopulations of cells with a single amplification.

Our study suggests that the assessment of co-amplification of *EGFR* and *CDK4* at the single-cell level by FISH could serve as a proxy of immune status. Standard immunofluorescence to assess immune infiltration is often difficult to quantify due to variation in staining intensity driven by technical issues and imaging modality and, most notably, due to limitations in accurate cell segmentation during image processing. Relying on a DNA-based nuclear signal of a FISH assay alleviates these challenges, as there is much less ambiguity about nuclear borders and speckle count compared with relative intensity of staining quantification. Quantification of FISH in thousands of cells could be easily implemented in a standard pathology lab setting, as our analyses were all done using open-source software and thresholding of the signal is easier to benchmark for FISH than for immunohistochemistry-based methods. Thus, the CNA status of subclonal populations of cells within a tumor could serve as a representation of tumor-intrinsic properties that can be linked to tumor microenvironmental status. Recent developments in spatial transcriptomics enabled the interrogation of the local patterns of immune infiltration in GBM.[24,57] However, a genomic landscape inferred from sparse transcriptomic data[59] does not yet allow for confident calling for CNAs at individual gene and single-cell levels. Combination of FISH and spatial transcriptomic copy-number inference at single-cell resolution may become a powerful tool in dissecting the interactions between clonally diverse populations of cancer cells and their microenvironment.

Limitations of the study

One limitation of this study is the relatively small size of our GBM cohort, which consisted of 17 tumors. The spatial analysis performed in this study describes the local diversity within a single biopsy and is thus not capturing location within the tumor or within the brain. Larger studies including multiregional biopsies and MRI-based location annotations will be required to capture features associated with macroscopic diversity in GBM.

Another limitation is that our amplification calls based on single-nuclei FISH signal counts in FFPE tumor tissue cannot be directly confirmed by single-cell sequencing methods. Whole-genome sequencing of FFPE samples at single-cell resolution[60] is challenging to perform at scale. Other methods, such as single-cell methylation profiling[61,62] or single-cell RNA-seq (scRNA-seq) provide sparse coverage, and thus smaller genomic region gains identified by FISH, such as *CDK4* in our study, may not be detected. Improvements in throughput and resolution of single-cell DNA profiling in archival specimen cohorts will be necessary to address these issues.

Future studies including a larger number of samples from clinical trials targeting CKD4, EGFR, or macrophages will shed more light on the utility of single-cell FISH signal quantification as a prognostic and predictive tool. Development of immunocompetent murine models recapitulating the genetic diversity of human tumors will also be needed to uncover the mechanisms driving the divergent co-evolution of the tumor and its microenvironment.

Our study provides evidence that the presence of genetically distinct subpopulations is associated with differences in the tumor microenvironment. This relationship opens novel avenues for utilizing established cytogenetic methods combined with

spatial profiling to improve our understanding of the complex cellular ecosystem of GBM.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Human tissue samples
  - Animals and tumor generation
  - Cell line generation
- METHOD DETAILS
  - DNA sequencing and analysis
  - GeoMx digital spatial profiling (DSP)
  - Fluorescence *in situ* hybridization (FISH)
  - STAR-FISH
  - Immunofluorescent staining
  - Flow cytometry analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - GeoMX proteomic analysis
  - Single-cell genotype analysis
  - Integrative analysis
  - Survival analysis
  - Single-cell transcriptomic data analysis

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2023.112235.

ACKNOWLEDGMENTS

We thank the members of the Janiszewska and Michor laboratories for their critical reading of this manuscript and useful discussions; Dr. Itay Tirosh and Dr. Julie Laffy from the Weizmann Institute of Science and Dr. Dan Landau from Weill Cornell Medicine, New York Genome Center, for help with manuscript revision; members of the UF Scripps Histology Core, Genomics Core, and Bioinformatics Core for their technical expertise; and Dr. Liang Zhang from Nanostring for help with GeoMX DSP experiment. This work was supported by the NIH K99/R00 CA201606 (M.J.), NIH 3R00 CA201606-05S1 (U.O.), the Center For Cancer Evolution (F.M.), The Helen Gurley Brown Presidential Initiative at Dana-Farber Cancer Institute (F.M. and D.E.), and start-up funds from the Scripps Research Institute (M.J.).

AUTHOR CONTRIBUTIONS

K.A.W., D.E., and S.C. are co-first authors, equally contributing to study conceptualization and execution. K.A.W. and D.E. performed GeoMX data analysis. D.E. performed single-cell genotype analysis, integrative dataset analysis, and survival data analysis. S.C. performed single-cell and whole-genome sequencing (WGS) analysis, with contributions from A.J. and J.G. K.A.W. performed validation of FISH, immunohistochemistry (IHC), fluorescence-activated cell sorting (FACS), and animal experiments. S.Y. performed TMA imaging and image analysis. U.O. perfromed TMA imaging and image analysis. R.S. performed STAR-FISH TERT promoter mutation-specific conditions. C.V. provided clinical samples used to construct the TMA and evaluated histopathological features of the TMA H&E staining. T.O.M. consulted on the statistical analysis. F.M. and M.J. supervised the study. All authors helped design the study and write the manuscript.

DECLARATION OF INTERESTS

M.J. is a member of a scientific advisory board at ResistanceBio (formerly Viosera Therapeutics).

INCLUSION AND DIVERSITY

We support inclusive, diverse, and equitable conduct of research. One or more of the authors of this paper self-identifies as an underrepresented ethnic minority in their field of research or within their geographical location. One or more of the authors of this paper self-identifies as a member of the LGBTQIA+ community. One or more of the authors of this paper self-identifies as a gender minority in their field of research.

Received: September 22, 2021
Revised: December 20, 2022
Accepted: February 23, 2023

REFERENCES

1. Wen, P.Y., Weller, M., Lee, E.Q., Alexander, B.M., Barnholtz-Sloan, J.S., Barthel, F.P., Batchelor, T.T., Bindra, R.S., Chang, S.M., Chiocca, E.A., et al. (2020). Glioblastoma in adults: a Society for Neuro-Oncology (SNO) and European Society of Neuro-Oncology (EANO) consensus review on current management and future directions. Neuro Oncol. *22*, 1073–1113. https://doi.org/10.1093/neuonc/noaa106.
2. Aldape, K., Brindle, K.M., Chesler, L., Chopra, R., Gajjar, A., Gilbert, M.R., Gottardo, N., Gutmann, D.H., Hargrave, D., Holland, E.C., et al. (2019). Challenges to curing primary brain tumours. Nat. Rev. Clin. Oncol. *16*, 509–520. https://doi.org/10.1038/s41571-019-0177-5.
3. Lee, J.-K., Wang, J., Sa, J.K., Ladewig, E., Lee, H.-O., Lee, I.-H., Kang, H.J., Rosenbloom, D.S., Camara, P.G., Liu, Z., et al. (2017). Spatiotemporal genomic architecture informs precision oncology in glioblastoma. Nat. Genet. *49*, 594–599. https://doi.org/10.1038/ng.3806.
4. Brennan, C.W., Verhaak, R.G.W., McKenna, A., Campos, B., Noushmehr, H., Salama, S.R., Zheng, S., Chakravarty, D., Sanborn, J.Z., Berman, S.H., et al. (2013). The somatic genomic landscape of glioblastoma. Cell *155*, 462–477. https://doi.org/10.1016/j.cell.2013.09.034.
5. Sottoriva, A., Spiteri, I., Piccirillo, S.G.M., Touloumis, A., Collins, V.P., Marioni, J.C., Curtis, C., Watts, C., and Tavaré, S. (2013). Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics. Proc. Natl. Acad. Sci. USA *110*, 4009–4014. https://doi.org/10.1073/pnas.1219747110.
6. Ozawa, T., Riester, M., Cheng, Y.-K., Huse, J.T., Squatrito, M., Helmy, K., Charles, N., Michor, F., and Holland, E.C. (2014). Most human non-GCIMP glioblastoma subtypes evolve from a common proneural-like precursor glioma. Cancer Cell *26*, 288–300. https://doi.org/10.1016/j.ccr.2014.06.005.
7. Rich, J.N., Reardon, D.A., Peery, T., Dowell, J.M., Quinn, J.A., Penne, K.L., Wikstrand, C.J., Van Duyn, L.B., Dancey, J.E., McLendon, R.E., et al. (2004). Phase II trial of gefitinib in recurrent glioblastoma. J. Clin. Oncol. *22*, 133–142. https://doi.org/10.1200/jco.2004.08.110.
8. Wen, P.Y., Yung, W.K.A., Lamborn, K.R., Dahia, P.L., Wang, Y., Peng, B., Abrey, L.E., Raizer, J., Cloughesy, T.F., Fink, K., et al. (2006). Phase I/II study of imatinib mesylate for recurrent malignant gliomas: north American Brain Tumor Consortium Study 99-08. Clin. Cancer Res. *12*, 4899–4907. https://doi.org/10.1158/1078-0432.ccr-06-0773.
9. Nicholson, J.G., and Fine, H.A. (2021). Diffuse glioma heterogeneity and its therapeutic implications. Cancer Discov. *11*, 575–590. https://doi.org/10.1158/2159-8290.cd-20-1474.

10. Snuderl, M., Fazlollahi, L., Le, L.P., Nitta, M., Zhelyazkova, B.H., Davidson, C.J., Akhavanfard, S., Cahill, D.P., Aldape, K.D., Betensky, R.A., et al. (2011). Mosaic amplification of multiple receptor tyrosine kinase genes in glioblastoma. Cancer Cell *20*, 810–817. https://doi.org/10.1016/j.ccr.2011.11.005.

11. Szerlip, N.J., Pedraza, A., Chakravarty, D., Azim, M., McGuire, J., Fang, Y., Ozawa, T., Holland, E.C., Huse, J.T., Jhanwar, S., et al. (2012). Intratumoral heterogeneity of receptor tyrosine kinases EGFR and PDGFRA amplification in glioblastoma defines subpopulations with distinct growth factor response. Proc. Natl. Acad. Sci. USA *109*, 3041–3046. https://doi.org/10.1073/pnas.1114033109.

12. Verhaak, R.G.W., Hoadley, K.A., Purdom, E., Wang, V., Qi, Y., Wilkerson, M.D., Miller, C.R., Ding, L., Golub, T., Mesirov, J.P., et al. (2010). Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. Cancer Cell *17*, 98–110. https://doi.org/10.1016/j.ccr.2009.12.020.

13. Wang, Q., Hu, B., Hu, X., Kim, H., Squatrito, M., Scarpace, L., deCarvalho, A.C., Lyu, S., Li, P., Li, Y., et al. (2017). Tumor evolution of glioma-intrinsic gene expression subtypes associates with immunological changes in the microenvironment. Cancer Cell *32*, 42–56.e6. https://doi.org/10.1016/j.ccell.2017.06.003.

14. Patel, A.P., Tirosh, I., Trombetta, J.J., Shalek, A.K., Gillespie, S.M., Wakimoto, H., Cahill, D.P., Nahed, B.V., Curry, W.T., Martuza, R.L., et al. (2014). Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. Science *344*, 1396–1401. https://doi.org/10.1126/science.1254257.

15. Neftel, C., Laffy, J., Filbin, M.G., Hara, T., Shore, M.E., Rahme, G.J., Richman, A.R., Silverbush, D., Shaw, M.L., Hebert, C.M., et al. (2019). An integrative model of cellular states, plasticity, and genetics for glioblastoma. Cell *178*, 835–849.e21. https://doi.org/10.1016/j.cell.2019.06.024.

16. Richards, L.M., Whitley, O.K.N., MacLeod, G., Cavalli, F.M.G., Coutinho, F.J., Jaramillo, J.E., Svergun, N., Riverin, M., Croucher, D.C., Kushida, M., et al. (2021). Gradient of Developmental and Injury Response transcriptional states defines functional vulnerabilities underpinning glioblastoma heterogeneity. Nat. Cancer *2*, 157–173. https://doi.org/10.1038/s43018-020-00154-9.

17. Garofano, L., Migliozzi, S., Oh, Y.T., D'Angelo, F., Najac, R.D., Ko, A., Frangaj, B., Caruso, F.P., Yu, K., Yuan, J., et al. (2021). Pathway-based classification of glioblastoma uncovers a mitochondrial subtype with therapeutic vulnerabilities. Nat. Cancer *2*, 141–156. https://doi.org/10.1038/s43018-020-00159-4.

18. O'Connor, J.P.B., Rose, C.J., Waterton, J.C., Carano, R.A.D., Parker, G.J.M., and Jackson, A. (2015). Imaging intratumor heterogeneity: role in therapy response, resistance, and clinical outcome. Clin. Cancer Res. *21*, 249–257. https://doi.org/10.1158/1078-0432.ccr-14-0990.

19. Tomaszewski, W., Sanchez-Perez, L., Gajewski, T.F., and Sampson, J.H. (2019). Brain tumor microenvironment and host state: implications for immunotherapy. Clin. Cancer Res. *25*, 4202–4210. https://doi.org/10.1158/1078-0432.ccr-18-1627.

20. Calabrese, C., Poppleton, H., Kocak, M., Hogg, T.L., Fuller, C., Hamner, B., Oh, E.Y., Gaber, M.W., Finklestein, D., Allen, M., et al. (2007). A perivascular niche for brain tumor stem cells. Cancer Cell *11*, 69–82. https://doi.org/10.1016/j.ccr.2006.11.020.

21. Janiszewska, M., Suvà, M.L., Riggi, N., Houtkooper, R.H., Auwerx, J., Clément-Schatlo, V., Radovanovic, I., Rheinbay, E., Provero, P., and Stamenkovic, I. (2012). Imp2 controls oxidative phosphorylation and is crucial for preserving glioblastoma cancer stem cells. Genes Dev. *26*, 1926–1944. https://doi.org/10.1101/gad.188292.112.

22. Lathia, J.D., Mack, S.C., Mulkearns-Hubert, E.E., Valentim, C.L.L., and Rich, J.N. (2015). Cancer stem cells in glioblastoma. Genes Dev. *29*, 1203–1217. https://doi.org/10.1101/gad.261982.115.

23. Prager, B.C., Xie, Q., Bao, S., and Rich, J.N. (2019). Cancer stem cells: the architects of the tumor ecosystem. Cell Stem Cell *24*, 41–53. https://doi.org/10.1016/j.stem.2018.12.009.

24. Ravi, V.M., Will, P., Kueckelhaus, J., Sun, N., Joseph, K., Salié, H., Vollmer, L., Kuliesiute, U., von Ehr, J., Benotmane, J.K., et al. (2022). Spatially resolved multi-omics deciphers bidirectional tumor-host interdependence in glioblastoma. Cancer Cell *40*, 639–655.e13. https://doi.org/10.1016/j.ccell.2022.05.009.

25. Rheinbay, E., Suvà, M.L., Gillespie, S.M., Wakimoto, H., Patel, A.P., Shahid, M., Oksuz, O., Rabkin, S.D., Martuza, R.L., Rivera, M.N., et al. (2013). An aberrant transcription factor network essential for Wnt signaling and stem cell maintenance in glioblastoma. Cell Rep. *3*, 1567–1579. https://doi.org/10.1016/j.celrep.2013.04.021.

26. Clevers, H., Loh, K.M., and Nusse, R. (2014). Stem cell signaling. An integral program for tissue renewal and regeneration: Wnt signaling and stem cell control. Science *346*, 1248012. https://doi.org/10.1126/science.1248012.

27. Huang, M., Zhang, D., Wu, J.Y., Xing, K., Yeo, E., Li, C., Zhang, L., Holland, E., Yao, L., Qin, L., et al. (2020). Wnt-mediated endothelial transformation into mesenchymal stem cell-like cells induces chemoresistance in glioblastoma. Sci. Transl. Med. *12*, eaay7522. https://doi.org/10.1126/scitranslmed.aay7522.

28. Barthel, F.P., Johnson, K.C., Varn, F.S., Moskalik, A.D., Tanner, G., Kocakavuk, E., Anderson, K.J., Abiola, O., Aldape, K., Alfaro, K.D., et al. (2019). Longitudinal molecular trajectories of diffuse glioma in adults. Nature *576*, 112–120. https://doi.org/10.1038/s41586-019-1775-1.

29. Lee, J.H., Lee, J.E., Kahng, J.Y., Kim, S.H., Park, J.S., Yoon, S.J., Um, J.-Y., Kim, W.K., Lee, J.-K., Park, J., et al. (2018). Human glioblastoma arises from subventricular zone cells with low-level driver mutations. Nature *560*, 243–247. https://doi.org/10.1038/s41586-018-0389-3.

30. Wang, Z., Sun, D., Chen, Y.-J., Xie, X., Shi, Y., Tabar, V., Brennan, C.W., Bale, T.A., Jayewickreme, C.D., Laks, D.R., et al. (2020). Cell lineage-based stratification for glioblastoma. Cancer Cell *38*, 366–379.e8. https://doi.org/10.1016/j.ccell.2020.06.003.

31. Weis, S.M., and Cheresh, D.A. (2011). Tumor angiogenesis: molecular pathways and therapeutic targets. Nat. Med. *17*, 1359–1370. https://doi.org/10.1038/nm.2537.

32. Pace, A., Dirven, L., Koekkoek, J.A.F., Golla, H., Fleming, J., Rudà, R., Marosi, C., Le Rhun, E., Grant, R., Oliver, K., et al. (2017). European Association for Neuro-Oncology (EANO) guidelines for palliative care in adults with glioma. Lancet Oncol. *18*, e330–e340. https://doi.org/10.1016/s1470-2045(17)30345-5.

33. Chiocca, E.A., Yu, J.S., Lukas, R.V., Solomon, I.H., Ligon, K.L., Nakashima, H., Triggs, D.A., Reardon, D.A., Wen, P., Stopa, B.M., et al. (2019). Regulatable interleukin-12 gene therapy in patients with recurrent high-grade glioma: results of a phase 1 trial. Sci. Transl. Med. *11*, eaaw5680. https://doi.org/10.1126/scitranslmed.aaw5680.

34. Keskin, D.B., Anandappa, A.J., Sun, J., Tirosh, I., Mathewson, N.D., Li, S., Oliveira, G., Giobbie-Hurder, A., Felt, K., Gjini, E., et al. (2019). Neoantigen vaccine generates intratumoral T cell responses in phase Ib glioblastoma trial. Nature *565*, 234–239. https://doi.org/10.1038/s41586-018-0792-9.

35. Magurran, A.E. (2005). Biological diversity. Curr. Biol. *15*, R116–R118. https://doi.org/10.1016/j.cub.2005.02.006.

36. Körber, V., Yang, J., Barah, P., Wu, Y., Stichel, D., Gu, Z., Fletcher, M.N.C., Jones, D., Hentschel, B., Lamszus, K., et al. (2019). Evolutionary trajectories of IDHWT glioblastomas reveal a common path of early tumorigenesis instigated years ahead of initial diagnosis. Cancer Cell *35*, 692–704.e12. https://doi.org/10.1016/j.ccell.2019.02.007.

37. Janiszewska, M., Liu, L., Almendro, V., Kuang, Y., Paweletz, C., Sakr, R.A., Weigelt, B., Hanker, A.B., Chandarlapaty, S., King, T.A., et al. (2015). In situ single-cell analysis identifies heterogeneity for PIK3CA mutation and HER2 amplification in HER2-positive breast cancer. Nat. Genet. *47*, 1212–1219. https://doi.org/10.1038/ng.3391.

38. Deshpande, V., Luebeck, J., Nguyen, N.-P.D., Bakhtiari, M., Turner, K.M., Schwab, R., Carter, H., Mischel, P.S., and Bafna, V. (2019). Exploring the landscape of focal amplifications in cancer using AmpliconArchitect. Nat. Commun. *10*, 392. https://doi.org/10.1038/s41467-018-08200-y.

39. Wu, S., Turner, K.M., Nguyen, N., Raviram, R., Erb, M., Santini, J., Luebeck, J., Rajkumar, U., Diao, Y., Li, B., et al. (2019). Circular ecDNA promotes accessible chromatin and high oncogene expression. Nature *575*, 699–703. https://doi.org/10.1038/s41586-019-1763-5.

40. Wang, L.-B., Karpova, A., Gritsenko, M.A., Kyle, J.E., Cao, S., Li, Y., Rykunov, D., Colaprico, A., Rothstein, J.H., Hong, R., et al. (2021). Proteogenomic and metabolomic characterization of human glioblastoma. Cancer Cell *39*, 509–528.e20. https://doi.org/10.1016/j.ccell.2021.01.006.

41. Liu, S., Zhang, C., Maimela, N.R., Yang, L., Zhang, Z., Ping, Y., Huang, L., and Zhang, Y. (2019). Molecular and clinical characterization of CD163 expression via large-scale analysis in glioma. OncoImmunology *8*, 1601478. https://doi.org/10.1080/2162402x.2019.1601478.

42. Zeng, F., Wang, K., Liu, X., and Zhao, Z. (2020). Comprehensive profiling identifies a novel signature with robust predictive value and reveals the potential drug resistance mechanism in glioma. Cell Commun. Signal. *18*, 2–13. https://doi.org/10.1186/s12964-019-0492-6.

43. Grzelak, C.A., Goddard, E.T., Lederer, E.E., Rajaram, K., Dai, J., Shor, R.E., Lim, A.R., Kim, J., Beronja, S., Funnell, A.P.W., and Ghajar, C.M. (2022). Elimination of fluorescent protein immunogenicity permits modeling of metastasis in immune-competent settings. Cancer Cell *40*, 1–2. https://doi.org/10.1016/j.ccell.2021.11.004.

44. Tirosh, I., Venteicher, A.S., Hebert, C., Escalante, L.E., Patel, A.P., Yizhak, K., Fisher, J.M., Rodman, C., Mount, C., Filbin, M.G., et al. (2016). Single-cell RNA-seq supports a developmental hierarchy in human oligodendroglioma. Nature *539*, 309–313. https://doi.org/10.1038/nature20123.

45. Turner, K.M., Deshpande, V., Beyter, D., Koga, T., Rusert, J., Lee, C., Li, B., Arden, K., Ren, B., Nathanson, D.A., et al. (2017). Extrachromosomal oncogene amplification drives tumour evolution and genetic heterogeneity. Nature *543*, 122–125. https://doi.org/10.1038/nature21356.

46. Morton, A.R., Dogan-Artun, N., Faber, Z.J., MacLeod, G., Bartels, C.F., Piazza, M.S., Allan, K.C., Mack, S.C., Wang, X., Gimple, R.C., et al. (2019). Functional enhancers shape extrachromosomal oncogene amplifications. Cell *179*, 1330–1341.e13. https://doi.org/10.1016/j.cell.2019.10.039.

47. Verhaak, R.G.W., Bafna, V., and Mischel, P.S. (2019). Extrachromosomal oncogene amplification in tumour pathogenesis and evolution. Nat. Rev. Cancer *19*, 283–288. https://doi.org/10.1038/s41568-019-0128-6.

48. Lange, J.T., Rose, J.C., Chen, C.Y., Pichugin, Y., Xie, L., Tang, J., Hung, K.L., Yost, K.E., Shi, Q., Erb, M.L., et al. (2022). The evolutionary dynamics of extrachromosomal DNA in human cancers. Nat. Genet. *54*, 1527–1533. https://doi.org/10.1038/s41588-022-01177-x.

49. Hung, K.L., Yost, K.E., Xie, L., Shi, Q., Helmsauer, K., Luebeck, J., Schöpflin, R., Lange, J.T., Chamorro González, R., Weiser, N.E., et al. (2021). ecDNA hubs drive cooperative intermolecular oncogene expression. Nature *600*, 731–736. https://doi.org/10.1038/s41586-021-04116-8.

50. Vogt, N., Lefèvre, S.H., Apiou, F., Dutrillaux, A.-M., Cör, A., Leuraud, P., Poupon, M.-F., Dutrillaux, B., Debatisse, M., and Malfoy, B. (2004). Molecular structure of double-minute chromosomes bearing amplified copies of the epidermal growth factor receptor gene in gliomas. Proc. Natl. Acad. Sci. USA *101*, 11368–11373. https://doi.org/10.1073/pnas.0402979101.

51. deCarvalho, A.C., Kim, H., Poisson, L.M., Winn, M.E., Mueller, C., Cherba, D., Koeman, J., Seth, S., Protopopov, A., Felicella, M., et al. (2018). Discordant inheritance of chromosomal and extrachromosomal DNA elements contributes to dynamic disease evolution in glioblastoma. Nat. Genet. *50*, 708–717. https://doi.org/10.1038/s41588-018-0105-0.

52. Johnson, K.C., Anderson, K.J., Courtois, E.T., Gujar, A.D., Barthel, F.P., Varn, F.S., Luo, D., Seignon, M., Yi, E., Kim, H., et al. (2021). Single-cell multimodal glioma analyses identify epigenetic regulators of cellular plasticity and environmental stress response. Nat. Genet. *53*, 1456–1468. https://doi.org/10.1038/s41588-021-00926-8.

53. Chen, T., Chen, J., Zhu, Y., Li, Y., Wang, Y., Chen, H., Wang, J., Li, X., Liu, Y., Li, B., et al. (2019). CD163, a novel therapeutic target, regulates the proliferation and stemness of glioma cells via casein kinase 2. Oncogene *38*, 1183–1199. https://doi.org/10.1038/s41388-018-0515-6.

54. Lisi, L., Ciotti, G.M.P., Braun, D., Kalinin, S., Currò, D., Dello Russo, C., Coli, A., Mangiola, A., Anile, C., Feinstein, D.L., and Navarra, P. (2017). Expression of iNOS, CD163 and ARG-1 taken as M1 and M2 markers of microglial polarization in human glioblastoma and the surrounding normal parenchyma. Neurosci. Lett. *645*, 106–112. https://doi.org/10.1016/j.neulet.2017.02.076.

55. Quail, D.F., and Joyce, J.A. (2017). The microenvironmental landscape of brain tumors. Cancer Cell *31*, 326–341. https://doi.org/10.1016/j.ccell.2017.02.009.

56. Luoto, S., Hermelo, I., Vuorinen, E.M., Hannus, P., Kesseli, J., Nykter, M., and Granberg, K.J. (2018). Computational characterization of suppressive immune microenvironments in glioblastoma. Cancer Res. *78*, 5574–5585. https://doi.org/10.1158/0008-5472.can-17-3714.

57. Varn, F.S., Johnson, K.C., Martinek, J., Huse, J.T., Nasrallah, M.P., Wesseling, P., Cooper, L.A.D., Malta, T.M., Wade, T.E., Sabedot, T.S., et al. (2022). Glioma progression is shaped by genetic evolution and microenvironment interactions. Cell *185*, 2184–2199.e16. https://doi.org/10.1016/j.cell.2022.04.038.

58. Rutledge, W.C., Kong, J., Gao, J., Gutman, D.A., Cooper, L.A.D., Appin, C., Park, Y., Scarpace, L., Mikkelsen, T., Cohen, M.L., et al. (2013). Tumor-infiltrating lymphocytes in glioblastoma are associated with specific genomic alterations and related to transcriptional class. Clin. Cancer Res. *19*, 4951–4960. https://doi.org/10.1158/1078-0432.ccr-13-0551.

59. Erickson, A., He, M., Berglund, E., Marklund, M., Mirzazadeh, R., Schultz, N., Kvastad, L., Andersson, A., Bergenstråhle, L., Bergenstråhle, J., et al. (2022). Spatially resolved clonal copy number alterations in benign and malignant tissue. Nature *608*, 360–367. https://doi.org/10.1038/s41586-022-05023-2.

60. Martelotto, L.G., Baslan, T., Kendall, J., Geyer, F.C., Burke, K.A., Spraggon, L., Piscuoglio, S., Chadalavada, K., Nanjangud, G., Ng, C.K.Y., et al. (2017). Whole-genome single-cell copy number profiling from formalin-fixed paraffin-embedded samples. Nat. Med. *23*, 376–385. https://doi.org/10.1038/nm.4279.

61. Gaiti, F., Chaligne, R., Gu, H., Brand, R.M., Kothen-Hill, S., Schulman, R.C., Grigorev, K., Risso, D., Kim, K.-T., Pastore, A., et al. (2019). Epigenetic evolution and lineage histories of chronic lymphocytic leukaemia. Nature *569*, 576–580. https://doi.org/10.1038/s41586-019-1198-z.

62. Chaligne, R., Gaiti, F., Silverbush, D., Schiffman, J.S., Weisman, H.R., Kluegel, L., Gritsch, S., Deochand, S.D., Gonzalez Castro, L.N., Richman, A.R., et al. (2021). Epigenetic encoding, heritability and plasticity of glioma transcriptional cell states. Nat. Genet. *53*, 1469–1479. https://doi.org/10.1038/s41588-021-00927-7.

63. Campeau, E., Ruhl, V.E., Rodier, F., Smith, C.L., Rahmberg, B.L., Fuss, J.O., Campisi, J., Yaswen, P., Cooper, P.K., and Kaufman, P.D. (2009). A versatile viral system for expression and depletion of proteins in mammalian cells. PLoS One *4*, e6529. https://doi.org/10.1371/journal.pone.0006529.

64. Enomoto, M., Bunge, M.B., and Tsoulfas, P. (2013). A multifunctional neurotrophin with reduced affinity to p75NTR enhances transplanted Schwann cell survival and axon growth after spinal cord injury. Exp. Neurol. *248*, 170–182. https://doi.org/10.1016/j.expneurol.2013.06.013.

65. Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nat. Methods *9*, 671–675. https://doi.org/10.1038/nmeth.2089.

66. Team, R.C. (2021). R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing).

67. Rossum, V., and Jr, D. (2009). Python 3 Reference Manual (Centrum voor Wiskunde en Informatica Amsterdam).

68. Janiszewska, M., Stein, S., Metzger Filho, O., Eng, J., Kingston, N.L., Harper, N.W., Rye, I.H., Alečković, M., Trinh, A., Murphy, K.C., et al. (2021). The impact of tumor epithelial and microenvironmental heterogeneity on treatment responses in HER2-positive breast cancer. JCI Insight *6*, e147617. https://doi.org/10.1172/jci.insight.147617.

69. Ochocka, N., Segit, P., Walentynowicz, K.A., Wojnicki, K., Cyranowski, S., Swatler, J., Mieczkowski, J., and Kaminska, B. (2021). Single-cell RNA sequencing reveals functional heterogeneity of glioma-associated brain macrophages. Nat. Commun. *12*, 1151. https://doi.org/10.1038/s41467-021-21407-w.

70. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M.A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. *20*, 1297–1303. https://doi.org/10.1101/gr.107524.110.

71. Talevich, E., Shain, A.H., Botton, T., and Bastian, B.C. (2016). CNVkit: genome-wide copy number detection and visualization from targeted DNA sequencing. PLoS Comput. Biol. *12*, e1004873. https://doi.org/10.1371/journal.pcbi.1004873.

72. Li, A., Bai, Q., Kong, H., Zhou, S., Lv, H., Zhong, S., Li, M., Bi, R., Zhou, X., and Yang, W. (2020). Impact of the updated 2018 American society of clinical oncology/college of American pathologists guideline for human epidermal growth factor receptor 2 testing in breast cancer. Arch. Pathol. Lab Med. *144*, 1097–1107. https://doi.org/10.5858/arpa.2019-0369-oa.

73. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. Cell *184*, 3573–3587.e29. https://doi.org/10.1016/j.cell.2021.04.048.

74. Murray, P.J., Allen, J.E., Biswas, S.K., Fisher, E.A., Gilroy, D.W., Goerdt, S., Gordon, S., Hamilton, J.A., Ivashkiv, L.B., Lawrence, T., et al. (2014). Macrophage activation and polarization: nomenclature and experimental guidelines. Immunity *41*, 14–20. https://doi.org/10.1016/j.immuni.2014.06.008.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit monoclonal anti-CD163 (clone EPR19518)	Abcam	Cat#ab182422; RRID:AB_2753196
Goat anti-rabbit Alexa Fluor 568	Invitrogen	Cat#A11011; RRID:AB_143157
Cd45-PE-Cy7	BD Biosciences	Cat#561868; RRID:AB_10893599
Cd11b-AlexaFluor700	BD Biosciences	Cat#557960; RRID:AB_396960
Ly6C-PerCP-Cy5.5	BD Biosciences	Cat#560525; RRID:AB_1727558
Ly6G-PacificBlue	Biolegend	Cat#127612; RRID:AB_2251161
anti-Egfr	Cell Signaling	Cat#2232S
anti-Cdk4	Abcam	Cat#ab137675
anti-GAPDH	Cell Signaling	Cat#8884S
Biological samples		
GBM FFPE blocks	Mount Sinai Medical Center, Miami, Florida, USA	N/A
Chemicals, peptides, and recombinant proteins		
dUTP	Abbott Molecular	Cat#02N32-050
Human COT1 DNA	Invitrogen	Cat#15279011
Vysis CEP hybridization buffer	Abbott Molecular	Cat#07J36-001
UltraPure 20x SSC	Invitrogen	Cat#15557044
NP-40	Sigma	Cat#13021
ProLong Gold Antifade mountant with DAPI	Invitrogen	Cat#P36931
dNTP	New England Bio Labs	Cat#N0446
7-deaza- 2′-deoxy-GTP	Roche	Cat#10988537001
Betaine	Alfa Aesar	Cat#J77507VCR
Platinum Taq Polymerase	Life Technologies	Cat#10966083
Proteinase K	ThermoFisher	Cat#AM2548
Antigen retrieval solution pH9	Dako	Cat#S2367
Goat serum	ThermoFisher	Cat#31873
TruStain FcX PLUS	Biolegend	Cat#156603
Staining Buffer	Biolegend	Cat#420201
Critical commercial assays		
QiaAmp DNA FFPE Tissue Kit	Qiagen	Cat#56404
AmpliSeq Cancer HotSpot Panel v2 kit	Illumina	Cat#20019161
GeoMx Digital Spatial Profiler (DSP) Protein panel - service	Nanostring	N/A
Nick Translation Kit	Abbott	Cat#07J00-001
DNA PCR-Free Library Prep kit	Illumina	Cat# 20041795
Deposited data		
Whole genome sequencing	This paper	phs003100.v1.p1
Single-cell RNA sequencing from GEO	Neftel et al.[15]	GSE131928
Single-cell RNA sequencing from Broad Institute Single-Cell Portal	Richards et al.[16]	SCP503
Experimental models: Cell lines		
U-87 MG	ATTC	Cat#HTB-14
DBTRG	DSMZ	Cat#ACC 359

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
HEK-293T	Sigma	Cat#12022001
GL261	DSMZ	Cat#ACC802
Oligonucleotides		
STAR-FISH hTERT F1: 5′CTATGGTTCCAGGCCCGTTC-3′	This paper	N/A
STAR-FISH hTERT R1: 5′GGCTCCCAGTGGATTCGC-3′	This paper	N/A
STAR-FISH hTERT F2: 5′TGTCGACGCAAAACCGGTTC CGGCCCAGCCCTTT-3′	This paper	N/A
STAR-FISH hTERT R2: 5′GCGATATGACGACGCGAAT ACCCACGTGCGCAGC-3′	This paper	N/A
Specific-to-allele PCR-FISH: +T + G+TCGACGCAAAACCGG+T+T + C (+ indicates LNA modified bases)	This paper	N/A
Recombinant DNA		
BAC clone RP11-339F13 (EGFR gene)	BACPAC Genomics	N/A
BAC clone RP11-231C18 (PDGFRA gene)	BACPAC Genomics	N/A
BAC clone RP11-571M6 (CDK4 gene)	BACPAC Genomics	N/A
pLenti-CMV-Puro	Campeau et al.[63]	Addgene Plasmid #17452
pLenti-CMV-Neo	Campeau et al.[63]	Addgene Plasmid #17392
pLV-eGFP	Enomoto et al.[64]	Addgene Plasmid #36083
pLV-mCherry	Gift from Pantelis Tsoulfas	Addgene Plasmid #36084
pLenti-CMV-Egfr-Puro	This paper	N/A
pLenti-CMV-Cdk4-Neo	This paper	N/A
Software and algorithms		
ImageJ	Schneider et al.[65]	https://imagej.nih.gov/ij/
R software	R Core Team[66]	https://www.r-project.org/
Python	Rossum et al.[67]	https://www.python.org/
Single cell FISH counting macro for ImageJ	Janiszewska et al.[68]	https://doi.org/10.1172/jci.insight.147617
FlowJo	Beckman Dickinson	https://www.flowjo.com/
AmpliconArchitect	Deshpande et al.[38]	https://github.com/nf-core/circdna.git
Code related to analyses	This paper	https://github.com/Michorlab/GBM_OR_immune https://doi.org/10.5281/zendo.7618100

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by lead contact, Michalina Janiszewska (mjaniszewska@ufl.edu).

Materials availability

There are restrictions to the availability of patient tissue samples due to IRB and small amount of tissue available for this study (archival blocks and microarray slide were mostly consumed during this study). Cell lines derived from GL261 are available upon request.

Data and code availability

- De-identified human whole genome sequencing data have been deposited at dbGaP and accession number is listed in the key resources table. They are available upon request if access is granted. To request access, contact dbGaP (dbGaP: https://dbgap.ncbi.nlm.nih.gov). In addition, public summary-level phenotype have been deposited at dbGaP (dbGaP: https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs003100.v1.p1) and are publicly available as of the date of publication. The accession numbers are also listed in the key resources table. Preprocessed GeoMx, FISH and

OncoPanel data are available as part of the Supplemental Tables. This paper also analyzes existing, publicly available data. These accession numbers for the datasets are listed in the key resources table.

- All original code has been deposited at https://github.com/Michorlab/GBM_OR_immune and is publicly available as of the date of publication. DOIs is listed in the key resources table.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Human tissue samples

All experiments with use of human tumor tissue were approved by Scripps Research IRB protocol #IRB-18-7209 and Mount Sinai Medical Center IRB. Formalin-fixed paraffin embedded (FFPE) tissue blocks were provided by Dr. Cristina Vincentelli, Mount Sinai Medical Center, under IRB exemption for discarded tissue. GBM pathology was confirmed for each block by a board-certified neuropathologist. The cohort was comprised of 20 cases (10 female, 10 male) of IDH wild-type GBM and one recurrence. Clinical details are shown in Table S1. Primary tumor samples were collected prior to treatment, with exception of a few cases receiving steroids before surgery. Tissue microarray (TMA) was constructed by manually selecting four distant areas within each block, cores of 1mm diameter were punched. H&E staining of the TMA was used to assess presence of hemorrhage, microvascular proliferations, and necrosis within each core, done by board certified pathologist, blinded to the other data derived from TMA. Two cores on H&E slide were partially folded during tissue sectioning, therefore histological assessment of these cores was not included in the analysis.

Animals and tumor generation

All animal experimental procedures were approved by the Institutional Animal Care and Use Committee (protocol #18–031). Seven-week-old male C57BL/6J mice (Jackson Laboratory) were injected intracranially with 80,000 GL261 cells in DMEM (Gibco). To generate OR^{low} and OR^{high} tumors, GL261 cells harboring EGFR, CDK4 or both oncogene overexpression were mixed as follows: ORlow 73% of E, 19% of C, and 8% of EC GL261 cells, ORhigh with 22% of E, 60% of C, and 18% of EC GL261 cells. The injections were performed using stereotaxic apparatus, to deliver the cells into right striatum (1mm forward and 2mm right of the anterior fontanelle, 3mm vertical depth). 21 days post implantations animals were sacrificed, perfused with cold sterile PBS, and brains were isolated into cold HBSS without ions and kept on ice. Dissociation of tumor and surrounding tissue was performed as previously described.[69]

Cell line generation

The paternal GL261 mouse glioma cell line (DSMZ, ACC802) was used to generate Egfr, Cdk4 and Egfr+Cdk4 overexpressing subpopulations by lentiviral transduction. Male sex of the cell line was confirmed by Y chromosome specific PCR. Murine *Egfr* or *Cdk4* was cloned into pLenti-CMV-Puro or pLenti-CMV-Neo plasmid (Addgene) and lentiviral particles were produced according to Life Technologies protocols. After viral transduction GL261 cells were selected with 1.5 μg/ml puromycin and/or 300ug/ml neomycin for 4 days and oncogene overexpression was confirmed by western blotting using anti-Egfr (Cell Signaling, 2232S, 1:1000), anti-Cdk4 (Abcam, ab137675, 1:5000), and anti-GAPDH (Cell Signaling, 8884S, 1:1000). Lentiviral particles for expression of fluorescent proteins were prepared in the same way as described above using pLV-eGFP and pLV-mCherry vectors (Addgene). Cells overexpressing eGFP and/or mCherry were sorted using BD FACS Aria III (BD Biosciences) and overexpression of the oncogenes was again confirmed by Western blot analysis.

All cell lines were cultured DMEM media (Gibco) with 10% fetal bovine serum (FBS, Gibco) and antibiotics (100 U/mL penicillin and 100 μg/mL streptomycin, PenStrep, Gibco) at 37 °C in a humidified atmosphere with 5% CO_2. Cells were routinely checked for mycoplasma contamination using Mycoplasma PCR Detection kit (Applied Biological Materials, G238).

METHOD DETAILS

DNA sequencing and analysis

DNA was extracted from 30um sections of each FFPE GBM block, before cores for TMA construction were removed, using QiaAmp DNA FFPE Tissue Kit (Qiagen) and libraries were prepared using AmpliSeq Cancer HotSpot Panel v2 kit (Illumina). This panel targets 2800 COSMIC mutations in 50 oncogenes and tumor-suppressor genes. The libraries were sequenced using 2 × 150bp MiSeq format. For analysis DNA Amplicon Workflow version 3.24.1.8 + master was used. Sequences were aligned to reference human genome hg19 using BWA-MEM Whole Genome Aligner version 0.7.9a-isis-1.0.2. Over 99.5% on-target aligned reads were reported for each sample, base mismatch was ~0.24%. Variant calling was done with Pisces Variant Caller 5.2.9.23, Illumina Annotation Engine 2.0.11-0-g7fb24a09, Bam Metrics v0.0.22, SAMtools 0.1.19-isis-1.0.3.

For the whole-genome sequencing DNA was extracted from three 4um sections of three OR^{high} and three OR^{low} tumors using QiaAmp DNA FFPE Tissue Kit (Qiagen) and libraries were prepared using DNA PCR-Free Prep kit (Illumina). The libraries were sequenced using 2 × 250bp NovaSeq6000 format.

Whole Genome sequencing analysis was conducted using the in-house Whole-genome sequencing (WGS) pipeline. WGS implements Gene Analysis Toolkit (GATK)[70] best practices and uses Sentieon's bwa-mem aligner to align the FASTQ to GDC hg38 reference genome. Sentieon's base quality recalibration is used on aligned BAM to generate recalibrated BAM. CNVkit[71] was used to determine the log2 copy number ratios from the recalibrated BAM. Further, chr12 and chr7 regions were plotted for all the samples to separate tumors into high grade and low grade. Amplicon Architect was run using the nf-core circdna (https://github.com/nf-core/circdna.git) pipeline to find the focal amplifications for *EGFR* and *CDK4* genes.

GeoMx digital spatial profiling (DSP)

FFPE TMA was profiled using the commercial GeoMX DSP platform (Nanostring). TMA was stained with three fluorescent visualization markers, GFAP (GBM cells and astrocytes), CD45 (immune cells), and αSMA (endothelial cells) and two panels of UV-cleavable oligo-labeled antibodies (Table S2). Stained slides were then digitally scanned, and 96 regions of interest (ROI) were selected based on the fluorescent channels, at least one ROI per TMA core. For TMA cores which displayed heterogeneity for the fluorescent markers, two ROIs were selected. Selected ROIs were UV-illuminated to release the conjugated oligos and quantified on the nCounter system (Nanostring).

Fluorescence *in situ* hybridization (FISH)

Bacterial artificial chromosome clones RP11-339F13 (*EGFR* gene), RP11-231C18 (*PDGFRA* gene), RP11-571M6 (*CDK4* gene) were obtained from BACPAC Genomics (formerly BACPAC Resources at Childern's Hospital Oakland Research Institute) and validated by PCR and FISH on xenografts with known amplification status (Figure S3). FISH probes were generated by Nick Translation (Abbott Molecular, 07J00-001) using fluorescent dUTPs (Abbott Molecular, 02N32-050). The FFPE TMA slide that was used in GeoMx DSP was de-mounted, washed in a series of increasing ethanol concentration solutions and incubated with hybridization solution containing 1:1:1 ratio of fluorescently labeled FISH probes, human COT1 DNA (Invitrogen, 15279011) and Vysis CEP hybridization buffer (Abbott Molecular, 07J36-001) at 74°C for 7min and then at 37°C overnight. The slide was then washed for 2min with 0.4x SSC/0.3% NP-40 at room temperature, 0.4x SSC/0.3% NP-40 at 74°C, 2x SSC/0.1% NP-40 at room temperature and 2x SSC at room temperature; buffers made with UltraPure 20x SSC (Invitrogen, 15557044) and NP-40 (Sigma, 13021). After wash with PBS and H_2O, the slide was dried and mounted with ProLong Gold Antifade mountant with DAPI (Invitrogen, P36931). Imaging was performed on Olympus FV3000 confocal microscope (Olympus), for each TMA core three z-stack images were taken and maximum projections were used to quantify signals from the three fluorescent channels in each individual nucleus (ImageJ macro available upon request).

STAR-FISH

The PCR conditions specific for *hTERT* C228T promoter mutation were optimized on DNA extracted from U87 (*hTERT* C228T mutant) and HEK-293T cells (*hTERT* C228 WT) with primers F1: 5′CTATGGTTCCAGGCCCGTTC-3′, R1: 5′GGCTCCCAGTGGATTCGC-3′ in first round and F2: 5′TGTCGACGCAAAACCGGTTCCGGCCCAGCCCTTT-3′, R2: 5′GCGATATGACGACGCGAATACCCACG TGCGCAGC-3′ in second round (IDT). 1st round PCR reactions were set with 2.5 mM $MgCl_2$, 200 μM dNTPs (with 3:1 mix of 7-deaza-dGTP:dGTP; dNTPs from New England Bio Labs, cat#N0446, 7-deaza- 2′-deoxy-GTP from Roche, cat#10988537001), 0.8 M Betaine (Alfa Aesar, cat#J77507VCR), 200 nM of each primer and 0.125 U of Platinum Taq Polymerase (Life Technologies, cat#10966083). After initial denaturation at 95°C for 1 min, 10 steps of 3 cycles each (95°C - 30 s, $T_{annealing}$ - 30 s, 72°C - 20 s) were performed, starting at $T_{annealing}$ = 70°C ending at $T_{annealing}$ = 61°C with $\Delta T_{annealing} = -1$°C between steps. Then 10 cycles were performed at $T_{annealing}$ = 60°C (95°C - 30 s, 60°C - 30 s, 72°C - 20 s), followed by final extension at 72°C for 1 min. PCR conditions for second round PCR were 2 mM $MgCl_2$, 200 μM dNTPs (with 3:1 7-deaza-dGTP:dGTP), 250 nM of each primer, with 0.2 U of Platinum Taq Polymerase. After initial denaturation at 95°C for 1 min, 18 steps of 3 cycles each (95°C - 30 s, $T_{annealing}$ - 30 s, 72°C - 20 s) were performed, starting at $T_{annealing}$ = 70°C ending at $T_{annealing}$ = 53°C with $\Delta T_{annealing} = -1$°C between steps. Then 15 cycles were performed at $T_{annealing}$ = 58°C (95°C - 30 s, 58°C - 30 s, 72°C - 20 s), followed by final extension at 72°C for 1 min.

Specific-to-allele PCR-FISH was performed as described previously.[37] Briefly, after deparaffinization the FFPE TMA slide was treated with Proteinase K (20μg/ml, ThermoFisher AM2548) and subjected to two rounds of *in situ* PCR with a mixture of primers specific to the *hTERT* C228T mutation. After *in situ* PCR, the slide was washed in a series of increasing ethanol concentration solutions and hybridization of a probe specific to the PCR product (/5FAM/+T + G+TCGACGCAAAACCGG+T+T + C (+ indicates LNA modified bases), custom made by Life Techonologies, working stock 25 μM) was performed at 74°C for 7min and continued overnight at 37°C. After post-hybridization washes, as in FISH protocol, the slide was mounted with ProLong Gold Antifade mountant with DAPI (Invitrogen, P36931). Imaging, 3 images per core, and image analysis was performed as for the FISH experiment (modified ImageJ macro from our previous work[68]). This quantification allowed for classification of each individual nucleus as WT or MUT for *hTERT* promoter mutation and downstream analysis was performed using R software.

Immunofluorescent staining

After deparaffinization, the FFPE sections were subject to antigen retrieval solution pH9 (Dako, S2367) for 20 min in a steamer. Blocking with 10% goat serum (ThermoFisher, 31,873) in PBST at room temperature was followed by incubation with primary antibody for CD163 (Abcam, ab182422, 1.4ug/ml working solution) at 4°C overnight. Next, the slides were washed 3 times with PBS and

incubated for 1h at room temperature with goat anti-rabbit Alexa Fluor 568 (ThermoFisher, Invitrogen A-1101, 5ug/ml working solution). After washes with PBS, the slides were mounted with with ProLong Gold Antifade mountant with DAPI (Invitrogen, P36931). Images were taken using Olympus FV3000 confocal microscope (Olympus) and image stitching was performed with CellSense (Olympus).

Flow cytometry analysis

Cells were washed with FACS buffer (2% FBS in PBS with 2mM EDTA), followed by a blocking step with anti-mouse Cd16/32 FC block (TruStain FcX PLUS, Biolegend, 156,603, 1:200) for 15 min on ice. Cells were then stained with titrated antibodies for Cd45-PE-Cy7 (BD Biosciences, 561,868, clone 30-F11, 1:400), Cd11b-AlexaFluor700 (BD Biosciences, 557,960, clone M1/70, 1:800), Ly6C-PerCP-Cy5.5 (BD Biosciences, 560,525, clone AL-21, 1:200), Ly6G-PacificBlue (Biolegend, 127612, 1:200) in Staining Buffer (Biolegend, 420201) at 4°C for 20min. Cells were washed and resuspended in 300ul of Staining Buffer. Samples were analyzed on BD LSR II cytometer (BD Biosciences). Gating and subpopulation analysis was performed using FlowJo software (BD Biosciences). Reported percentages are given as the percentage of a parental gate, singlets discrimination was based on FSC and SSC scatters.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical details for each experiment can be found in the respective figure legends.

GeoMX proteomic analysis

The data were normalized to positive controls from the External RNA Controls Consortium (ERCC) and IgG isotype controls per ROI to account for differences in hybridization efficiency, background signal, cellularity, and size of ROIs. Tumor-level (core-level) protein expression was calculated from ROI-level protein expression as a weighted (by ROI nucleus number) average over all ROIs in a tumor (core) and min-max rescaled to a [0,1] range for each protein to enable global comparison. Tumor-level expression profiles were used in calculating Spearman correlations between proteins and performing hierarchical clustering on the dataset (Figure 2C). ROI-level expression was used for establishing the relative inter-tumor to intra-tumor variability in protein expression via the Kruskal-Wallis H test (Python SciPy) (Figure 2D). Here, each tumor constitutes a group, with within-group samples constituting all ROIs within a tumor. All p values except where indicated are at the $p < 0.01$ significance level after FDR adjustment for multiple comparisons. Figure 2C indicates distinct protein clusters, derived from hierarchical clustering. Correlations between pairs of clusters was computed as the correlation between scores comprising the sum of normalized protein levels for all proteins in each cluster.

Single-cell genotype analysis

Cores for which no GeoMX data was available were removed from the analysis of both amplifications and *TERT* mutations due to tissue quality concerns. Cells were considered to be amplified in a gene g if at least 6 copies of g were recorded in that nucleus location, analogous to the *HER2* FISH scoring guidelines of the American Society of Clinical Oncology/College of American Pathologists.[72] Coordinates for which 50 or greater copies of any gene were recorded were removed from the dataset as these likely represent overlapping cells. The Shannon diversity index for genotypic diversity in each individual image i (Figure 3D) was computed as $H_i = -\sum_{G \in G} R_{G,i} \log R_{G,i}$, where $R_{G,i}$ represents the proportion (if nonzero) of cells of genotype G in image i out of the set G of four genotypes: E, C, EC, and N/O. Tumor-level genotype proportions were computed by aggregating (adding) data from all images in a particular tumor. Spearman correlations between proportions (Figures 3E and S3C) were computed based on tumor-level proportions using the R Hmisc package, with p values FDR-adjusted for multiple comparisons. Best-fit lines shown in Figure 3E were obtained with an inverse (y = a/x + b) curve fit (Python SciPy) (top panels; negative Spearman coefficient) and linear fits (Python NumPy polyfit) (bottom panels; positive Spearman coefficient). Computation of minimum relative distances (Figure 4G) was done by calculating in each image i, for each cell c of genotype G, the Euclidean distance $d^{(c)}_{G,G}$ to the nearest cell of the genotype G' and the Euclidean distance $d^{(c)}_{G,G'}$ to the nearest cell of any genotype. The average minimum relative distance for genotype G and genotype G' (where we also consider $G = G'$) in image i is then given by

$$\langle \mathrm{R_{G,i}} \rangle = \langle \frac{\mathrm{d^{(c)}_{G,G}}}{\mathrm{d^{(c)}_{G,G'}}} \rangle_{c \in C_{i,G}}$$

where $C_{i,G}$ is the set of all cells of genotype G in image i. Computation of clustering properties proceeded by implementing k-means clustering (Python scikit-learn) in each image i for all cells $c \in C_{i,G}$. The optimal number of clusters (from 2 to 5) was selected by finding the highest Silhouette score. The total sum of squares (TSS) was computed as Euclidean distances between all $c \in C_{i,G}$ and the mean of c, $c \in C_{i,G}$. The between cluster sum of squares (BCSS) was then computed as the difference between the TSS and the within cluster sum of squares, as given by the inertia property of the clusters obtained. BCSS/TSS represents the ratio of the between cluster sum of squares to the total sum of squares and provides a measure for the tightness of clustering of each genotype, with a higher ratio indicating tighter clustering.

The Shannon diversity index for copy number diversity in each individual image (Figure 4F) was calculated on a tumor level, after aggregating all the images in that tumor. The frequency of each copy number in the relevant cell type (e.g., E cells) was divided by the total number of cells of that type, and then the sum over those was taken. For a cell type \omega that would be \sum_{\omega copy number}\frac{frequency of copy number}{total number of \omega cells}.

Integrative analysis

The odds ratio of *EGFR* and *CDK4* co-amplification (EC genotype) is given by the odds of a cell being co-amplified given that it is amplified in *EGFR* relative to the odds of a cell being co-amplified given that it is not amplified in *EGFR*, $OR = (N_{EC}/N_E)/(N_C/N_{N/O})$, where N is the number of cells with a given genotype. The OR is calculated as follows:

The odds ratio of a CDK4 amplification in a cell if EGFR is amplified in a cell is given by

$$O_A = \frac{N_{EC}/(N_{EC}+N_E)}{1 - N_{EC}/(N_{EC}+N_E)}$$

And the odds of a CDK4 amplification inna cell if EGFR is not amplified in a cell is given by

$$O_N = \frac{N_C/(N_C+N_{N/O})}{1 - N_C/(N_C+N_{N/O})}$$

So that the odds ratio of *EGFR* and *CDK4* co-amplification is given by

$$OR = \frac{O_A}{O_N} = \frac{\frac{N_{EC}/(N_{EC}+N_E)}{1-N_{EC}/(N_{EC}+N_E)}}{\frac{N_C/(N_C+N_{N/O})}{1-N_C/(N_C+N_{N/O})}} = \frac{\frac{N_{EC}/(N_{EC}+N_E)}{N_E/(N_{EC}+N_E)}}{\frac{N_C/(N_C+N_{N/O})}{N_{N/O}/(N_C+N_{N/O})}} = \frac{N_{EC}/N_E}{N_C/N_{N/O}}$$

The OR was computed in each tumor separately by combining data from all images in the tumor (Figure 4A). For subsequent analysis, tumors were divided based on their computed OR into tertiles representing low-OR (OR^{low}), middle-OR, and high-OR (OR^{high}) tumors. To determine statistical significance of distributional differences between the top and bottom tertiles in across the various covariates tested, Mann-Whitney tests (Python SciPy) were employed. Associated proteomic analysis (Figures 5A and 5B) employed tumor-level expression data, processed as described above.

Survival analysis

Separate analyses were carried out on clinical and proteomics data from Wang et al.[40] (Figure 5D) and on clinical and mRNA expression levels (U133 microarray; Figure 5E) TCGA Firehose Legacy dataset. In each case, clinical data was combined with CD163 expression data to fit a Cox proportional hazard model (Breslow's method), adjusted for patient age and sex, and compute a 95% CI. IDH-mutant tumors were excluded from the analysis. The Python lifelines package was used to fit the model, obtain Kaplan-Meier curves, and perform a log rank test.

Single-cell transcriptomic data analysis

Single cell CNA were inferred from two different scRNAseq datasets: Richards et al.[16] and Neftel et al.,[15] using the inferCNV R package v.1.3. Data from both studies was downloaded from the Single Cell Portal maintained by the Broad Institute, together with sample metadata, including cell type assignment, as follows: normalized expression values log2(TPM/10 + 1) for Neftel et al.[15] and UMI counts for Richards et al.[16]

inferCNV was ran separately for the malignant cells of each patient with a standard window size of 100, and with the following parameters: HMM = true, HMM_type = 'i3', cluster_by_groups = TRUE, denoise = TRUE, analysis_mode = subcluster. For Neftel et al.,[15] cutoff = 1 was employed (this data was generated with the SmartSeq2 technology), while for Richards et al.,[16] cutoff = 0.1 was used (data generated with 10X Genomics). inferCNV employs a 3-state Hidden Markov Model, inferring neutral regions, amplifications, and deletions. Reference cell types were set to Macrophages, Oligodendrocytes and T cells on the Neftel et al.,[15] and to Immune and Normal Brain on the data in Richards et al.[16] Both datasets were assessed for the copy number status of EGFR and CDK4.

On the Richards et al.[16] data, a Seurat object[73] was created, with keeping only genes expressed in at least 10 cells, and only cells that express at least 200 features. Further, cells for which more than 20% of the transcriptome was encoded by mitochondrial genes were removed (43,177 cells analyzed). The data was preprocessed following the standard Seurat scRNAseq workflow, including normalizing with SCTransform, running PCA, constructing the nearest neighbor graph from the top 30 PC components, and further constructing the 2D UMAP representation. Cell types were assigned as non-immune, M1 and M2 macrophages, and T cells, using heatmaps with expression of canonical markers,[74] and expression scores of M1 and M2 macrophages (Tables S6 and S7). This amounted to 16,629 cells classified as non-immune, 1,484 cells as T-cells, 18,134 as M1, and 6,920 as M2 macrophages. This assignment matched very well with the provided annotation classifying cells into immune and non-immune.