nature communications

Article

https://doi.org/10.1038/s41467-025-59805-z

Resolving spatial subclonal genomic heterogeneity of loss of heterozygosity and extrachromosomal DNA in gliomas

Received: 15 August 2023

Accepted: 23 April 2025

Published online: 13 June 2025

Check for updates

Michelle G. Webb[1], Frances Chow[2], Carmel G. McCullough[1], Bohan Zhang[1], John J. Y. Lee[3], Rania Bassiouni[1], Norman E. Garrett III[2], Kyle Hurth[4], John D. Carpten[1,3], Gabriel Zada[2] ✉ & David W. Craig[1,3] ✉

Mapping the spatial organization of DNA-level somatic copy number changes in tumors can provide insight to understanding higher-level molecular and cellular processes that drive pathogenesis. We describe an integrated framework of spatial transcriptomics, tumor/normal DNA sequencing, and bulk RNA sequencing to identify shared and distinct characteristics of an initial cohort of eleven gliomas of varied pathology and a replication cohort of six high-grade glioblastomas. We identify focally amplified extrachromosomal DNA (ecDNA) in four of the eleven initial gliomas, with subclonal tumor heterogeneity in two *EGFR*-amplified grade IV glioblastomas. In a *TP53*-mutated glioblastoma, we detect a subclone with *EGFR* amplification on ecDNA coupled to chromosome 17 loss of heterozygosity. To validate subclonal somatic aneuploidy and copy number alterations associated with ecDNA double minutes, we examine the replication cohort, identifying *MDM2/MDM4* ecDNA subclones in two glioblastomas. The spatial heterogeneity of *EGFR* and p53 inactivation underscores the role of ecDNA in enabling rapid oncogene amplification and enhancing tumor adaptability under selective pressure.

Gliomas are a type of brain tumor originating from glial cells. They are classified into major types based on cell of origin, histological characteristics, and molecular alterations, which play a vital role in patient prognosis and treatment. Key subtypes include oligodendrogliomas, astrocytomas, diffuse midline gliomas, and glioblastomas. Each can have markedly different growth rates and varying responses to treatment, where subtypes such as oligodendrogliomas are slow-growing and more responsive to treatment, while others, such as astrocytomas and glioblastomas, are aggressive and difficult to treat[1].

Hallmark molecular alterations that define glioma subtypes include somatic mutations, histone gene alterations, chromosomal anomalies, and gene amplifications, and have evolved into defining roles in the histological diagnosis and classification of gliomas. For instance, oligodendrogliomas originate around cells that create and maintain myelin sheaths around nerve cells called oligodendrocytes and are molecularly characterized by co-deletion of chromosome 1p/19q and mutations in either *IDH1* or *IDH2*[2]. Astrocytomas originate from star-shaped glial cells and are further classified by their aggressiveness and typically *ATRX* and *TP53* mutations, contributing to tumor migration, evasion of apoptosis, and proliferation[3]. Although considered a subtype of astrocytoma, the term glioblastoma refers to astrocytic *IDH* wild-type tumors that include loss of function mutations in the tumor suppressor gene *PTEN* on chromosome 10, and frequent activation of PI3K/AKT/mTOR signaling, promoting cell growth and survival. *EGFR* mutations are characteristic of *IDH* wild-type glioblastomas, where overexpression promotes cell proliferation and

[1]Department of Translational Genomics, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA. [2]Department of Neurological Surgery, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA. [3]Department of Integrative Translational Sciences, City of Hope, Duarte, CA, USA. [4]Department of Pathology, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA.
✉e-mail: Gabriel.Zada@med.usc.edu; dacraig@coh.org

survival[4]. Diffuse midline gliomas are an additional subtype of astrocytomas, which include frequent *TP53* mutations, *PIK3CA* mutations, and signature *H3F3A* K27M mutations.

While understanding the underlying spatial heterogeneity of gliomas is essential for understanding the molecular processes that drive tumor pathogenesis, our knowledge of the spatial context of these tumor-driving molecular events has been limited. For example, gliomas are characterized by their diffuse nature, and tumors such as glioblastomas are fundamentally heterogeneous, where the tumor core is often molecularly distinct from tissue at the infiltrating edge and influenced by the surrounding tumor microenvironment. A recent approach to characterize invasive, non-enhancing high-grade gliomas with magnetic resonance imaging profiling identified heterogeneity of tumor populations[5]. Single-cell studies of gliomas can classify tumor subtypes, deconvolute microenvironment gene expression, and survey cellular plasticity, yet lack the spatial context provided by more traditional but targeted histopathological methods[6].

Recently, emerging spatial transcriptomics (ST) methods such as 10X Genomics Visium Spatial Gene Expression allow the assessment of cellular RNA expression globally and within the context of neighboring cells and structures, providing new insights into complex cell microenvironments and spatially distinct transcriptional patterns. In gliomas, Kim et al. applied an integrated approach of spatial profiling, single-cell RNA sequencing, and neuropathology analyses to characterize and resolve the molecular heterogeneity of the glioblastoma microenvironment[7]. Separately, Ren et al. used short and long-read spatial analysis to describe the expression and splicing of different isoforms in glioma niches and how this affected transcriptional signatures[8]. Ravi et al. applied multi-omic technologies to spatially resolve regional copy number alterations and exclusive transcriptional signatures while providing new software for spatial data[9,10]. Likewise, Jain et al. investigated cancer-associated fibroblasts and glioma stem cell interactions, identifying spatially relevant marker expression and cell type localizations[11]. More recently, multiple studies have examined high-grade glioblastomas by integrating single-cell RNA sequencing, spatial transcriptomics, and other multi-omic methods to better understand the relation between tissue architecture and invasion of the tumor microenvironment[12–15].

One aspect that has only been explored to a limited extent is how somatic DNA alterations relate to spatial biology within and across different types of gliomas. Across glioma subtypes, chromosome-level changes, or aneuploidy, are found in characteristic patterns. As stated previously, a hallmark of oligodendrogliomas is the co-deletion of chromosomes 1p and 19q due to a non-balanced translocation. Loss of 9p, where the tumor suppressor *CDKN2A* and adjacent *CDKN2B* are located, is a prognostic marker for astrocytomas and oligodendrogliomas consistent with significantly lower survival rates[16,17]. The most frequent numerical chromosome alterations in *IDH* wild-type glioblastoma include whole chromosome 7 gain and chromosome 10 loss, which often co-occur. Copy number alterations in gliomas can potentially be resolved using ST, allowing for the identification of somatic DNA-level copy number drivers accompanying gene expression changes across different tumor regions.

A hallmark of many gliomas, including glioblastoma, is the presence of highly amplified oncogenic driver genes on extrachromosomal DNA (ecDNA), often in the form of double-minute chromosomes (DMs)[18–20]. Early studies using fluorescence in situ hybridization (FISH) identified ecDNA elements containing amplified oncogenes, such as *EGFR* in gliomas, supporting the idea that ecDNA drives rapid oncogene amplification outside the chromosomal environment, contributing to glioma pathogenesis, genetic instability, and intratumoral heterogeneity[21]. Recent research has reinforced the role of ecDNA in driving tumor progression, with amplifications in oncogenes such as *MDM2*, *MDM4*, and *CDK4*[22–24]. While spatial heterogeneity is recognized as a fundamental aspect of glioma progression, traditional FISH methods have limited spatial ability to resolve concomitant expression and genomic variation. New spatial and single-cell genome-wide methodologies offer a deeper understanding of how ecDNA-driven oncogenic events co-occur and influence tumor development[23,25].

In this work, we describe a spatial transcriptomic analysis of several major classifications of gliomas, including five glioblastomas, two astrocytomas, three oligodendrogliomas, and one diffuse midline glioma. Considering the pronounced molecular heterogeneity and aneuploidy inherent to gliomas, the spatial analysis of gene expression and somatic variations can provide insights into tumor biology and pathogenesis. We utilize tumor-normal exome sequencing combined with spatial transcriptomics data to gain additional insights into chromosomal and ecDNA level changes. Generating expression-based analysis and bulked sequencing data of both tumor DNA and RNA, we examine RNA allele-specific expression patterns to detect regions of loss of heterozygosity (LOH) and copy number alteration using Bayesian modeling and hidden Markov modeling (HMM) algorithms. Integrating this approach across both bulk and spatial data on this varied group of gliomas, we discover subclonal loss of heterozygosity events supported by clinical reports and confirmed by analysis of the DNA sequences of tumor and normal tissue samples both in an initial and replication cohort.

Results

Initial sample cohort selection and rationale in glioma cohort A

There were two primary objectives in sample selection. The first was to conduct a comparative analysis of different glioma types, aiming to characterize the inherent variability across gliomas. Overall, sections included five World Health Organization (WHO) grade 4 glioblastomas, one WHO grade 2 astrocytoma, one WHO grade 4 astrocytoma, three WHO grade 2 oligodendrogliomas, and one WHO grade 4 diffuse midline glioma. Additional considerations were to have at least one subtype with different WHO grades and at least two sections from the same individual at recurrence. This aspect of the study design allowed us to understand global molecular drivers of variation in the tumor microenvironment, identify genes distinguishing subtypes, and inform future studies focused on variation within subtypes. The second component included an in-depth analysis of five high-grade glioblastomas. Glioblastomas are known to be highly heterogeneous tumors with different cell populations and often harbor subclonal populations with distinct genetic and transcriptomic profiles.

Spatial transcriptomics combined with genomic data can help identify subclonal populations and their spatial distribution, informing our understanding of tumor evolution. Our goal was to characterize the spatial heterogeneity in terms of genomic somatic DNA events and the transcriptomic alterations driving pathogenesis. Samples underwent comprehensive molecular analysis, utilizing Visium spatial transcriptomics, exome sequencing of tumor-derived DNA, and RNA-sequencing of tumor-derived RNA.

Clinical and molecular features of samples in the glioma cohort A

For clarity, samples throughout the text are referred to by their pathology and a unique numeric identifier. Figure 1 shows clinical and molecular features for all samples, where molecular features have also been confirmed by clinical labs and histopathology. Histological grade, microsatellite instability score, and disease-relevant mutation status were abstracted from clinical records, shown in Fig. 1. Of note, Glioblastoma A1 and Glioblastoma A2 were WHO grade 4, *IDH* wild-type, *EGFR*vIII positive tumors from the same patient. We note that Astrocytoma A1 was also characterized as an oligoastrocytoma in some clinical histopathology, indicating a mixed lineage, but we utilize the updated secondary astrocytoma label throughout the text. These tumors were included in the study design to observe any possible

spatial genomic changes resulting from chemotherapy treatment and disease recurrence.

Hallmark molecular alterations from bulk exome sequencing in the glioma cohort A

Initial characterization of all samples was performed by bulk exome sequencing of paired tumor and germline DNA, where tumor DNA was extracted from separately cut curls than that used for spatial transcriptomics. The genomic tumor DNA was sequenced to an average target coverage of 675x, with more than 95% of the targeted bases over 50x. Target enrichment was employed to assist with copy number analysis. For each sample in the study, the germline exome was sequenced to a mean target depth of 85x +/−20 from whole blood. Hallmark mutations and chromosomal copy number changes were determined through somatic variant calling and copy number analysis of paired tumor/germline exome sequenced data. The sequencing metrics described in this section are annotated by the sample in Supplementary Data 1. Hallmark mutations and copy number profiles specific to each sample are summarized in Fig. 1.

All glioblastoma samples were *IDH* wild-type with a hallmark gain of chromosome 7, loss of chromosome 10, and gain of chromosome 19. All except Glioblastoma A3 have chromosome 20 gain. Glioblastoma A2 harbored a *TP53* mutation and an *EGFR* mutation (c.754 C > T; R252C), which were also detected in the recurrence from the same individual (Glioblastoma A1). Additional *EGFR* mutations were detected in Glioblastoma A3 (c.866 C > T; A289V) and Glioblastoma A5 (c.787 A > C; T263P). All glioblastomas harbored a deletion of *CDKN2A*. *PDGFRA* was focally amplified in Glioblastoma A4, while *EGFR* was focally amplified in Glioblastoma A1, Glioblastoma A2, and Glioblastoma A5. As we subsequently discuss, these high-level focal amplifications are typically extrachromosomal DMs.

Three oligodendrogliomas were sequenced, two with an *IDH2* mutation and one with an *IDH1* mutation. Each of these showed the characteristic chromosome 1p and chromosome 19q deletions. Oligodendroglioma A1 had an additional *TP53* mutation, and Oligodendroglioma A3 had *PIK3CA* and *RB1* mutations. Both astrocytomas were *IDH1* mutants with *TP53* mutations. Finally, the Diffuse Midline Glioma A1 had a characteristic *H3F3A* (K27M) mutation, an *ATRX* stop mutation (R2111*), and a *PDGFRA* focal amplification containing a *PDGFRA* (c.1027 C > T; P343S) mutation.

Distinguishing and shared spatial transcriptomic features across glioma cohort A

The sample cohort includes gliomas of varied spatial complexity and genomic profiles. We applied Visium Spatial Gene Expression (10X Genomics) to all 11 fresh-frozen tissue samples, where key spatial experiment metrics are included in Supplementary Data 2. Sequencing data was analyzed with the 10X spaceranger pipelines.

We conducted both individual-level clustering (Supplementary Fig. 1) and inter-tumor clustering (Fig. 2). Each approach showed strengths and weaknesses in different contexts in their ability to reflect intra- and inter-tumor spatial biology. Specifically, individual-level spatial clustering was used to identify unique heterogenous features within a specific tumor to emphasize intratumor variability. Conversely, individual clustering does not reflect the global features of a tumor in comparison to others. For example, Glioblastoma A1 and Glioblastoma A2 are derived from the same patient at two different time points. The latter shows clusters driven by *VEGFA*, whereas the former has uniformly high coverage of *VEGFA*, making this feature less evident. Moreover, the fact that each tumor is clustered and normalized within its own space makes comparison across tumors difficult. We show individual-level clustering within the supplementary materials, focusing on integrated clustering.

Figure 2 shows the inter-tumor spatial biology by integrating the ST data from the discovery glioma set of samples. We used Seurat SCTransform normalization and reciprocal PCA integration methods, ensuring accurate inter-sample clustering. By aligning the datasets in a shared low-dimensional space, reciprocal principal component analysis preserved biological variation and allowed for reliable identification of consistent spot-level populations across samples. One prominent feature of ST compared to single-cell sequencing is that we see partially distinct clusters, reflecting that a 55-μm Visium spot overlaps multiple cells and cell types. In the primary integrated analysis, nine clusters were identified, representing shared cell populations present in samples at varying amounts (Fig. 2b, c), and mapping of integrated clusters on individual samples revealed a structured tissue architecture (Fig. 2d). Marker genes, or statistically significant overexpressed genes, were identified for each integrated cluster to better characterize the underlying biology (Fig. 2e). Apart from the primary analysis, it is important to note different features are evident at different resolutions. Two additional clustering analyses with smaller and larger resolutions are demonstrated in Supplementary Fig. 2.

The largest first blue cluster contained mixed neuronal lineages but was dominated by oligodendrocyte marker genes within the three oligodendroglioma samples. Co-expression in small clusters overexpressing *MAG*, *MOG*, and *MBP* (also in oligodendrocyte precursor cells) and oligodendrocyte precursor markers such as *OLIG1* and *OLIG2*. Additionally, the diffuse nature of cluster 1 may be drive by lower unique molecular identifier (UMI) counts in areas that include apoptotic and highly necrotic regions, as evidenced by the presence of

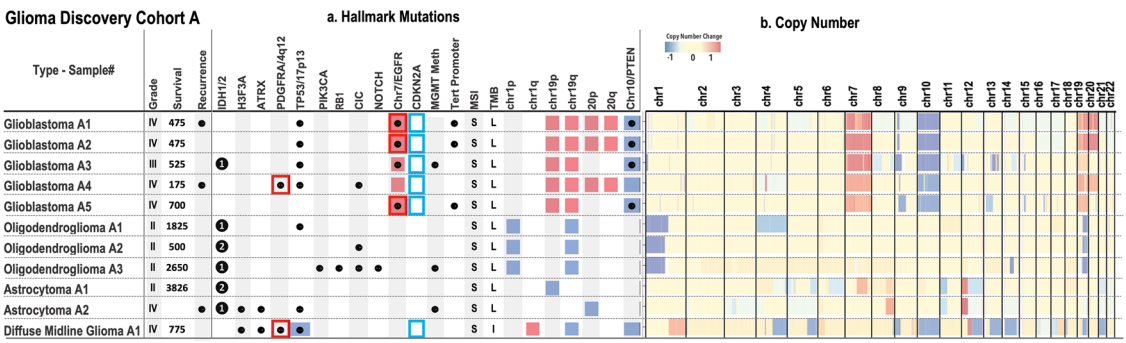

Fig. 1 | Overview of somatic alterations in the discovery glioma cohort A. The initial dataset of 11 fresh frozen glioma tumors were analyzed with spatial transcriptomics, bulk tumor and normal exome sequencing, and bulk RNA sequencing. The sample set included five glioblastomas, three oligodendrogliomas, two astrocytomas, and a diffuse midline glioma. Glioblastoma A1 was a recurrence of Glioblastoma A2. **a** Hallmark mutations identified through exome sequencing and somatic variant calling analysis are displayed with their mutation status. Point mutations are shown as •, *IDH1* mutations are 1, *IDH2* are 2. Microsatellite instability and tumor mutation burden are listed as MSI and TMB, respectively. **b** Copy number changes were identified through exome copy number variation analysis. Blue and red shading indicated chromosome level gains/losses, and red/blue borders indicate focal amplifications.

https://doi.org/10.1038/s41467-025-59805-z

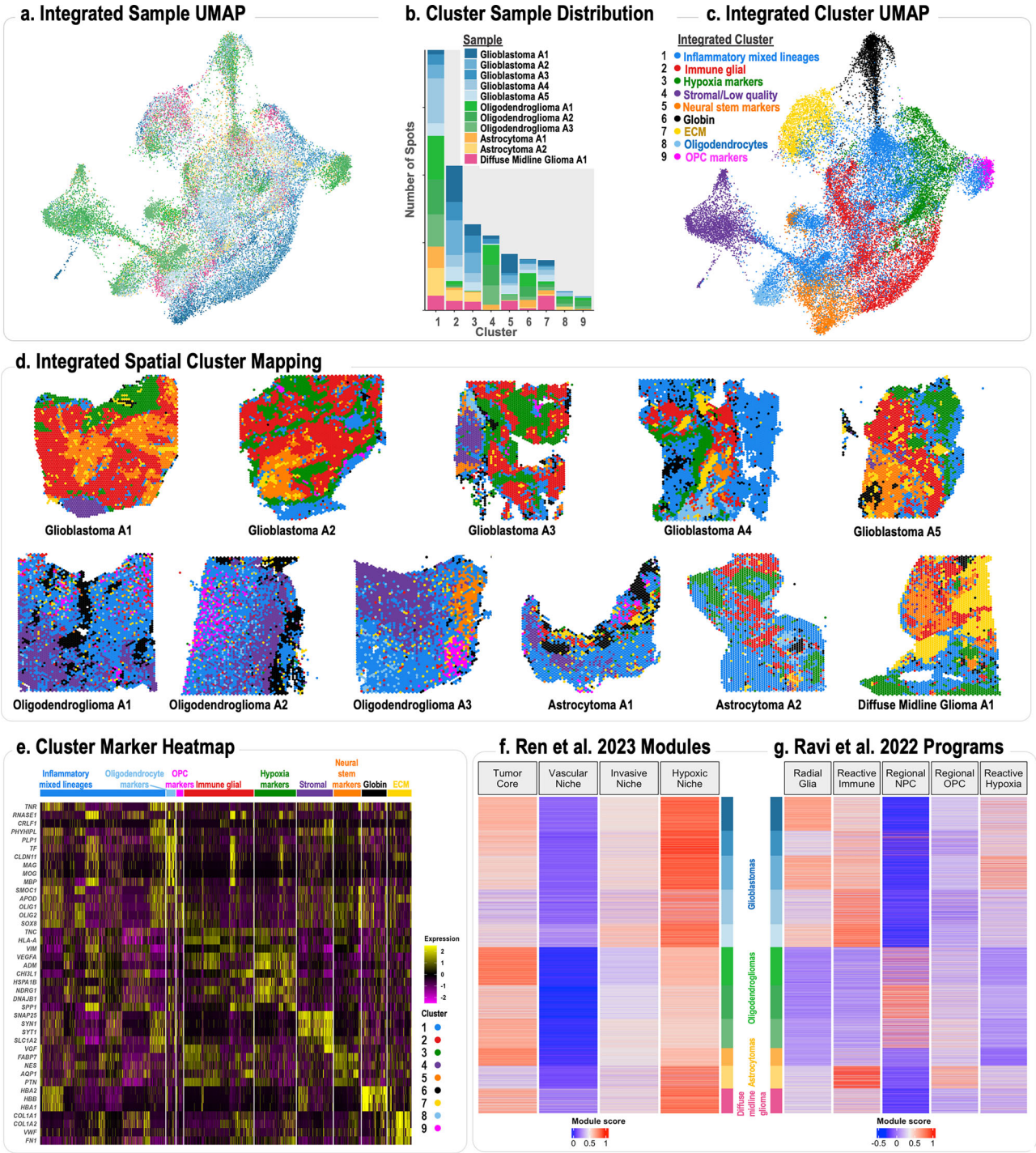

Fig. 2 | Glioma cohort A integration. Eleven gliomas of varied grading and molecular phenotype were analyzed by spatial transcriptomics were integrated bioinformatically with Seurat SCTransform normalization and reciprocal PCA workflow. **a** Uniform manifold approximation and projection (UMAP) of integrated dataset, highlighted by sample. **b** Stacked bar chart of the distribution of samples within the integrated clusters. **c** UMAP of the integrated dataset is color-coded by cluster. **d** Spatial maps of integrated cluster assignments for each sample. **e** Expression heatmap of top marker genes per cluster. **f** Heatmap of integrated data module scores for glioma niche-specific transcriptional modules previously described by Ren et al. 2023. The gene sets include tumor core, vascular niche, invasive niche, and hypoxic niche. **g** Heatmap of integrated data module scores for transcriptional programs previously described by Ravi et al. 2022. Subgroups include Radial Glia, Reactive Immune, Regional Neural Progenitor-like Cells, Regional Oligodendrocyte Progenitor-like Cells, and Reactive Hypoxia.

HBB and *HBA1* (globin markers) in black cluster 6. However, cluster 6 also contains other markers, such as *PLA2G2A* in astrocytoma cells or *MBP*, a marker of oligodendrocytes. Larger mixed lineage cluster 1 areas are found in the oligodendroglioma and astrocytoma tumors.

The oligodendrogliomas shared two additional clusters, 8 and 9. Both clusters contained marker for oligodendrocytes, with cluster 9 specifically exhibiting spatial separation of genes typically attributed to oligodendrocyte precursor cells, e.g., *OLIG1*, *OLIG2*, and *SMOC1*. Cluster 8 showed an overrepresentation of oligodendrocyte cell markers involved in myelination including *PLP1*, *TF*, *CLDN11*, *MAG*, *MOG*, and *MBP*. Notably, *MOG* encodes the myelin oligodendrocyte glycoprotein and *MBP* encodes the myelin basic protein. *PLP1* is an

https://doi.org/10.1038/s41467-025-59805-z

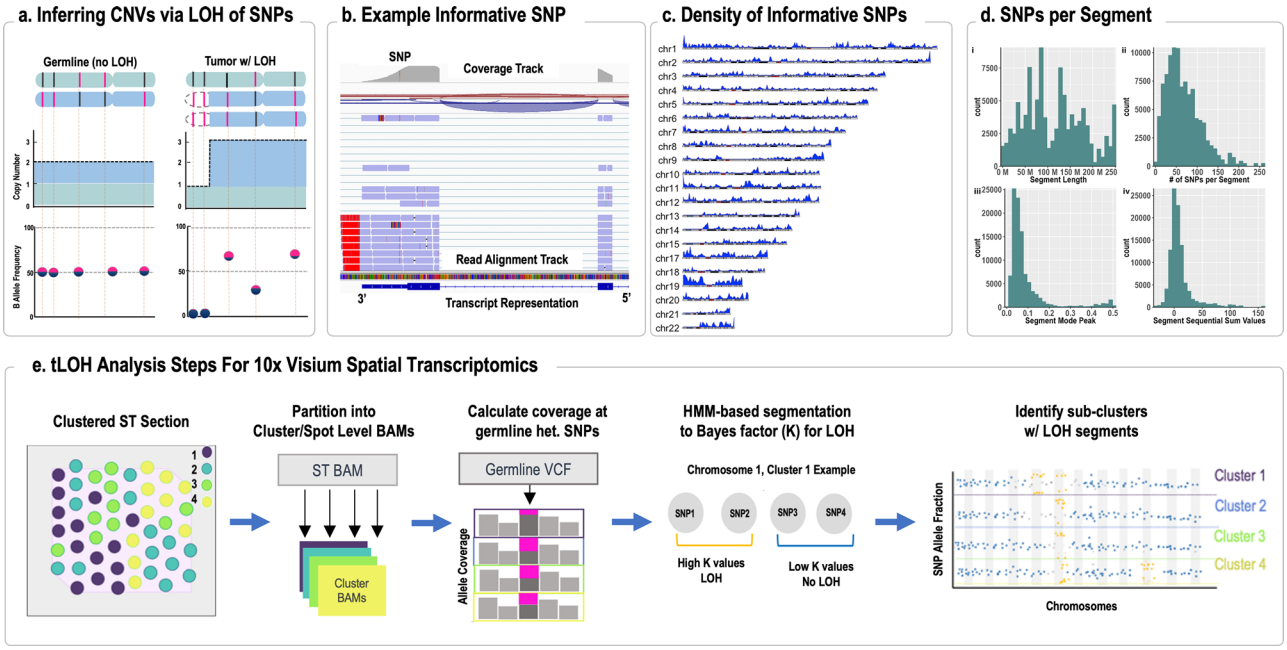

Fig. 3 | Loss of heterozygosity analysis workflow. a In our LOH analysis, the allele fraction of known heterozygous SNPs is used to infer underlying copy number changes at the DNA level. We compare germline and tumor scenarios, focusing on 5 SNPs to illustrate key principles. The germline example displays a copy number of 2 and B allele frequencies at 0.5 across all SNPs. In contrast, the tumor displays partial p-arm deletion (white) and chromosome duplication (blue). The copy number state varies from 1 in the deleted p-arm region to 3 in the duplication region. B-allele frequencies at each SNP vary from 0, 0.3, and 0.6. **b** Integrative Genomics Viewer example of aligned spatial transcriptomic sequencing reads. The top coverage track represents the total reads aligned, with an indication of a SNP highlighted in red. The reference transcript at the bottom is in the 5′ to 3′ orientation, as indicated by arrows. **c** The SNP density plot shows the location and relative quantity of unique SNPs across the glioma dataset. The chromosomes are listed on the y-axis, with an ideogram below density measurements. **d** Analysis of **i** length of defined segments, **ii** number of unique SNPs per segment, **iii** segment mode peak values, and **iv** segment sequential sum of log10(K) values. **e** The analysis workflow of LOH identification begins with a spatial sample analyzed with the 10X Genomics spaceranger software. A sample BAM is split into cluster-specific BAMs. Read coverage is calculated at predetermined heterozygous SNP positions, filtered by strict criteria. Bayes factor K values are calculated at each SNP. A hidden Markov model independently evaluates each chromosome for each cluster and assigns regions with state determinations. Metrics across segments are evaluated, and a final assignment of heterozygous, LOH, or undefined is determined. SNP allele fractions are plotted in different panels for each cluster, and points are color-coded by state.

oligodendrocyte marker involved in myelin production in the central nervous system, and *CLDN11* is critical for proper oligodendrocyte functioning.

Three clusters were shared primarily in the gliomas with astrocytic or glial cells: cluster 2, cluster 3, and cluster 5. Clusters 2 and 3 were found in glioblastoma, astrocytomas, and the diffuse midline glioma sample and largely absent from the oligodendroglioma samples, consistent with astrocyte lineage, in particular. As seen in Fig. 3e, there is considerable overlap between the two clusters with several hypoxia markers, including *VEGFA*. However, cluster 2 shows strong evidence for immune signaling (including *CD74, HLA-C, HLA-B, and HLA-A*) and multiple markers for radial glia, including *TAP1*, *GBP1*, and *ISG15*. Cluster 3 shows stronger signals of reactive hypoxia or glia, including *TAP1*, *BGP1*, and *ISG15*. Cluster 3 shows stronger signals of reactive hypoxia or mesenchymal cells through overexpression of *DNAJB1*, *CHI3L1*, *SPP1*, and *HSPA1B*. Cluster 5 shared similarity to both clusters exhibiting overexpression of neural stem cell markers, including *NES* and *AQP1*, though they are absent from the diffuse midline glioma.

Collagen genes (e.g., *COL1A1*, *COL1A2*, and *COL3A1*) and markers of extracellular matrix remodeling, including *VWF*, *IGFBP7*, and *FN1*, characterized cluster 7. While this cluster was seen in all samples, enrichment of these genes was most pronounced in the diffuse midline glioma. Cluster 6 was primarily driven by vascularization and dominated by hemoglobin *HBA1*, *HBA2*, and *HBB* genes. Finally, the composition of cluster 4 corresponded with stromal, or in some cases, necrotic tissue.

Comparison across tumor types

We characterize expression profiles across different tumor types, considering some limitations due to sample variability. Among the glioblastoma samples, 4 out of 5 exhibited highly amplified focal regions, with 3 samples showing *EGFR* amplification and 1 with *PDGFRA* amplification. In contrast, *PDGFRA* amplification was also found in the diffuse midline glioma sample, which also shared some molecular similarities with glioblastoma, such as *CDKN2A* deletion and chromosome 10 deletions, impacting *PTEN*, found in all glioblastoma samples. While this study only examined one diffuse midline glioma, we provide the results and data for inclusion in future studies. As highlighted earlier, enrichment of a cluster associated with collagen genes was most pronounced in this sample, along with the absence of a cluster dominated by neural stem cell marker expression.

As expected, all oligodendrogliomas showed the characteristic 1p/19q co-deletion. Genes associated with glial hypoxia, such as *VEGFA*, *ADM*, *HSPA1B*, *SPP1*, and *VIM*, were overexpressed in the glioblastoma samples. Additionally, classical mesenchymal genes like *CHI3L1* and *CD44* were distinctly overexpressed in glioblastoma, further highlighting its mesenchymal subtype characteristics. We observed the highest expression of astrocytic markers such as *GFAP* in glioblastoma samples, with moderate expression in astrocytomas. However, Oligodendroglioma A3 uniquely showed elevated *GFAP* expression, despite also displaying hallmark oligodendroglioma mutations, including the 1p loss. The oligodendroglioma samples were enriched for oligodendrocyte progenitor cell markers, including *OLIG1*, *OLIG2*, and *SMOC*.

Our initial focus was on identifying common features across the 11 gliomas, but we also explored grouping or pooling smaller subsets, such as by recurrence status, as seen with Glioblastoma A1 and Glioblastoma A4. However, the limited sample size prevented definitive molecular drivers, making meaningful comparisons difficult without larger sample sizes. Nevertheless, a consistent feature across the

gliomas was the presence of focal amplifications of *EGFR* or *PDGFRA* in 5 of the 11 samples. These features are discussed in more detail later and informed the selection of the Glioblastoma cohort B validation set.

Comparison across tumor types with published transcriptomic models

To further characterize transcriptomic profiles shared by the sample cohort, we utilized published gene sets describing glioma niche transcriptional programs. The modules of Tumor Core, Vascular, Invasive, and Hypoxic Niches defined by Ren et al. were derived from geographically weighted regression-based correlation of marker genes from spatially informed clusters across a dataset of tumors. A module score was calculated for each spatial spot, and the scores were analyzed in the aggregate for each sample and module (Fig. 2f). Overall, all samples scored highly in the Tumor Core and Hypoxic Niche modules, with lower scores for Vascular Niche and Invasive Niche. Oligodendroglioma A1 and Astrocytoma A1 displayed the highest Tumor Core module scores. High Hypoxic Niche scores correlated with our marker gene analysis, showing high expression of genes related to hypoxia pathways, such as *VEGFA*, *HSPA1B*, and *NDRG1*.

Loss of heterozygosity analysis from ST RNA-seq

Bulk sequencing identified major clonal events throughout the tumor, obtained from neighboring sections. Germline heterozygous SNPs from sequencing of paired blood are proxies for inferring subclonal loss of heterozygosity by assessing the allelic balance of RNA-expressed SNPs between the reference and alternative (B-allele) from copy number gains and losses. LOH analysis begins by splitting a spatial binary alignment map (BAM) file into cluster-specific BAMs, each containing only primary read alignments for a single cluster. To infer copy number changes in glioma ST data, we examine sequence-level data at heterozygous single nucleotide (SNP) sites to assess the ratio of allelic expression, where A-allele indicated the reference allele and B-allele indicates the non-reference allele. The B-allele frequency, therefore, equals the number of sequenced reads for B divided by all the reads counted at the SNP position. The number of heterozygous SNPs (AB) is expected to have a B-allele frequency of 50 percent at a normal diploid copy state. Copy number changes in the tumor impact this frequency. For example, a three-copy state with duplication of the B-allele (ABB) would have a B-allele frequency of 67 percent (Fig. 3a). If the frequency deviates from 50 percent towards 0 and 100 percent with adequate evidence of coverage, loss of heterozygosity can be determined. Since LOH can range from a focal event to stretching across an entire chromosome, more inferences must be drawn from survey data from contiguous SNPs. To account for different levels of expression, Bayes factor values are calculated for each SNP position, providing a measure of a model of 50/50 heterozygosity verses LOH models of allelic imbalance.

Segmentation analysis of LOH by HMM

Only one SNP or gene may be subject to various allele-specific RNA expression. A 2-state hidden Markov model segments multiple consecutive SNPs into genomic regions of similarity. Consecutive log10 scaled Bayes factor K values are added in each Markov-assigned region to create an aggregate value. Median region K values and these aggregate values are evaluated against a theoretically defined threshold to assign a label of LOH, heterozygous, or undefined. The three main metrics influencing final state determinations are the mode of segment allele fractions, segment log10(K) sum adjusted values, and median segment K values. We evaluated segments to obtain measures of significance and the extend of effect by analysis of segment modes. Measurements within a segment with the top 25% of read coverage were selected to calculate the mode allelic fraction of segments, and the shift in allele fraction was also calculated. Specifically, kernel smoothing is applied to obtain the mode peak to identify regions with strong shifts away from 0.5. The segment log10(K) sum value is the sum of all a segment's informative log10(K) values. Values between −0.5 and 0.5 are omitted from this calculation as they do not provide evidence for Model 1 (LOH) or Model 2 (heterozygosity). It should be noted that we did not fully explore the use of alternative means of developing initial starting parameters, including bulk sequencing data, and that further refinement may be possible beyond the scope of this work.

Segment log10(K) sum values greater than 20 were considered for state determination of LOH if the median is above 0.5. The median K value of each segment provides essential weight to the final state determination, as sum values can sometimes lead to false positives if there are many SNPs with minimal positive K values in a large region. Another metric to aid in state determination is the region threshold score. We calculate a threshold score by multiplying 85% of SNPs of 0.5, the smallest acceptable K value for evidence of loss of heterozygosity. If the segment sequential sum exceeds this threshold and has a mode peak greater than 0.1, but the median may not pass the cutoff, then the state is still assigned LOH. We evaluated these approaches with the dataset and benchmarked the results against known copy number events determined from exome sequencing and clinical reports. Specifically, Fig. 3c shows overall informative SNP density across each chromosome and key output metrics describing segment lengths. An overview of the output from our LOH identification method across each sample is in Supplementary Fig. 3, and bulk copy number analysis of the exome sequencing is provided in Supplementary Fig. 4.

We evaluated multiple methods to determine the best approach for this analysis, including assessing known LOH frequencies of glioma data for prior information. However, this yielded segments with very few observations, which were not robust enough for broader conclusions. Additional testing will be needed for application of this method on alternative tumor types. For example, parameter tuning would be necessary to analyze tumors with homologous recombination defects with higher rates of LOH. Adjusting the initial state and transition probabilities in this package impacts the model's segment output. Parameters such as the alpha/beta values in the Bayes factor calculation and transition probabilities substantially affect the final annotations. In our tests, adjusting the transition states proved unsuitable for the model. If default HMM parameters were not used for the predictions, the output would likely show more significant variability in the segmentation across datasets for each sample. The state annotations 'LOH', 'HET', or 'Undefined' are determined by cumulative metrics across segments with predefined cutoffs. If any default parameters were adjusted, we recommend revisiting and re-evaluating these cutoffs to maintain consistency and reliability in state labeling.

One specific example of subclonal genomic alteration involves loss of heterozygosity in Glioblastoma A1. LOH was observed on chromosome 17 in the *EGFR*-amplified cluster, and by phasing heterozygous SNPs, we identified 34 informative markers. This subclonal LOH pattern was consistent with findings from bulk exome sequencing (Supplementary Fig. 4). Notably, subclonal heterogeneity was also detected in Glioblastoma A5. Here, k-means clustering ($k = 3$) revealed several clusters, with LOH detected in chromosomes 7, 10, 13, 15, and 19 in the cluster with high *EGFR* expression. Re-clustering better defined these subclonal alterations and highlighted their role in tumor progression.

Correlation of LOH with tumor and stromal content

In studying subclonal heterogeneity, separating tumor and non-tumor regions is an essential initial consideration to reduce false positives or negatives with defining subclones. We employed ESTIMATE to assess tumor purity and stromal enrichment across samples transcriptionally to determine our samples' tumor and stromal content[26]. Spatial application of ESTIMATE has previously been

shown to correlate closely to a pathologist's annotation of tumor regions[27]. We examined the correlation of tLOH results with tumor content in the sample Diffuse Midline Glioma 1 (Fig. 4). Graph clustering analysis of this sample produced 7 clusters (Fig. 4a). Clusters showed varied stromal scores, with cluster 1 containing the highest score (Fig. 4b). A spatial map of the ESTIMATE determined tumor purity scores for Diffuse Midline Glioma A1 shows the highest tumor content in clusters 4 and 6 (Fig. 4c). ESTIMATE stromal scores for this sample show high stromal content in regions overlapping cluster 1 (Fig. 4d). The spatial gene expression of *PDGFRA*, a commonly amplified marker gene of diffuse midline gliomas, corresponds to the tumor content described in clusters 4 and 6 (Fig. 4e). This is supported by the spatial LOH analysis, which reflects an absence of LOH in a stromal cluster while LOH is present in tumor clusters (Fig. 4f, g). It should be noted that the potential use of LOH to identify stromal regions is not necessarily specific to Diffuse Midline Glioma A1, only that LOH was most evident in this sample.

Examination of LOH within astrocytoma, oligodendroglioma, and diffuse midline glioma samples

In the astrocytoma, oligodendroglioma, and diffuse midline glioma samples, the most prominent regional LOH differences aligning with clusters are the delineation between stromal and tumor cells, which is most evident in the diffuse midline glioma, as discussed above. We do not observe consistently strong LOH that definitively demonstrates subclonal genomic heterogeneity of biologically relevant subclones within the astrocytoma and oligodendroglioma samples. While we detect positive signals and allelic shifts within clusters (Supplementary Fig. 3), we typically observe SNPs with high coverage at approximately 50% allele fraction, as well as near 0% and 100% allele fractions. These observations suggest a possible mixture of cell types within the 55-μm Visium spatial spots. Furthermore, when examining the bulk sequencing data, we find inconsistent agreement with hallmark features such as chromosome 1p gain and associated LOH in oligodendrogliomas or LOH around chromosome 19p. Collectively, these results suggest a mixture of stromal and glioma cells, consistent with the fact that midline gliomas, oligodendrogliomas, and astrocytomas often exhibit less genomic instability compared to glioblastomas[28]. These tumors may possess more homogeneous cell populations with fewer subclonal variations, resulting in weaker or inconsistent LOH signals.

Focal amplifications of *EGFR*+/*PDGFRA*+ in high-grade glioblastomas

We observe the strongest evidence for subclonal LOH within the glioblastoma samples, particularly in Glioblastoma A1 and A5. This evidence includes both allelic shifts across clusters and supporting validation from bulk sequencing data, as well as targeted sequencing. In these samples, the focal amplification of ecDNA in these glioblastomas may contribute to genomic instability, leading to pronounced subclonal heterogeneity characterized by reproducible instability of key chromosomes observed in ST. Overall, high-level oncogenic amplification of receptor tyrosine kinases, specifically *PDGFRA* and *EGFR*, was observed in 5 of the 11 discovery samples. *EGFR* amplification was detected in Glioblastoma A1, A2, and A5, while *PDGFRA* amplification was identified in Glioblastoma A4 and Midline Glioma A1, highlighting the key role of the genes and genomic structures in driving tumor progression. High-level focal amplifications within gliomas are often found in the form of ecDNA structures known as DMs, which are small, circular DNA fragments that exist independently of chromosomal DNA. Studies have shown that DMs are particularly common in high-grade gliomas, where they harbor amplified oncogenes such as *EGFR*, enhancing tumor progression and resistance to therapy.

Characterizing subclonal spatial heterogeneous ecDNA in Glioblastoma A1 and A5

Mechanistically, ecDNA DM structures are important in cancer biology due to their distinct characteristics, including the lack of centromeres and telomeres, which allow them to replicate independently and rapidly amplify the copy number of key oncogenes outside normal cell cycle checkpoints. These ecDNA DMs allow cancer cells to dynamically reprogram their genome at the DNA level in rapid response to environmental and therapeutic pressures. Within glioblastomas, these structures were previously studied in karyotyping and are known to be a key driver of spatial genomic heterogeneity.

Detecting subclonal genomic alterations, particularly those driven by ecDNA, is crucial for understanding glioma progression. Figure 5 shows two examples of *EGFR*-amplified glioblastoma ecDNA fragments derived from copy number analysis of tumor sequencing depth. The high depth of exome sequencing was sufficient to detect the reads spanning the breakpoints of the ecDNA (see Fig. 5, panel v), confirming the circular structure of the DM elements in both Glioblastoma A1 and Glioblastoma A5. In Glioblastoma A1, two distinct but overlapping DMs were identified, both including the *EGFR* gene. One DM was larger at 3.7 Mb, while the other was smaller at 500 kb. Notably, the sample from Glioblastoma A1 is a recurrence of the same individual's tumor as Glioblastoma A2, and we observe only the smaller *EGFR* ecDNA in the primary tumor. This mutational evolution suggests that the larger DM in the recurrent tumor may have been selected as a response to therapy.

To further understand their architecture, we examined the amplified region's boundaries, particularly the breakpoints. Shown in Fig. 5 (right), there is a single DM spanning two distinct regions of Glioblastoma A5, where breakpoint analysis allows us to identify how the ecDNA DM is spliced together. The DM splices a 700 kb region from chromosome 7p22.3 containing the tumor suppressor *EIF3B* to a 900 kb region containing *EGFR*. *EIF3B* is a core subunit of the eukaryotic translation initiation factor 3 complex, which is crucial for sustaining the increased metabolic and proliferative demands of rapidly dividing tumor cells. In glioblastoma, *EIF3B* amplification may enhance the translational of specific oncogenes or survival factors, promoting aggressive tumor behavior. The co-amplification of *EGFR* and *EIF3B* suggests a synergistic mechanism in glioblastoma progression. On one side, *EGFR* amplification drives oncogenic signaling pathways that increase cell proliferation and survival, while on the other side, *EIF3B* amplification enhances the cell's capacity for synthesis of proteins involved in cell cycle control, apoptosis, and stress responses, all of which are critical for tumor progression and evolution.

Spatial mapping correlated expression of ecDNA-encoded genes

With knowledge of focally amplified genes, we observed a tight correlation of their gene expression that frequently led to distinct clusters from single-sample clustering. For instance, in Glioblastoma A1, unsupervised clustering of gene expression data revealed a subpopulation over-expressing *EGFR*, *LANCL2*, and *VOPP1*, which all map to the amplified ecDNA and show tight spot-to-spot correlation in log2 normalized transformed counts (Fig. 5a, panel vi). Importantly, subclonality was independently validated by a Foundation One clinical report, which confirmed *EGFR* amplification and detected an *EGFR*vIII R252C subclonal variant on a separately analyzed tissue section. The tight correlation of genes encoded within the same DM ecDNA is highly evident on Glioblastoma A5, where break-point analysis shows eight consecutive genes on chromosome 7p22.3 and four consecutive genes on the 7p11 *EGFR* locus form a block of high pair-wise Pearson correlation of normalized counts across the amplification (Fig. 5a, panel vi). It is important to highlight that the key to this analysis is the prior finding of the amplified regions by exome sequencing since typical ecDNA DMs contain too few genes for accurate de novo discovery/inference of copy number variance.

https://doi.org/10.1038/s41467-025-59805-z

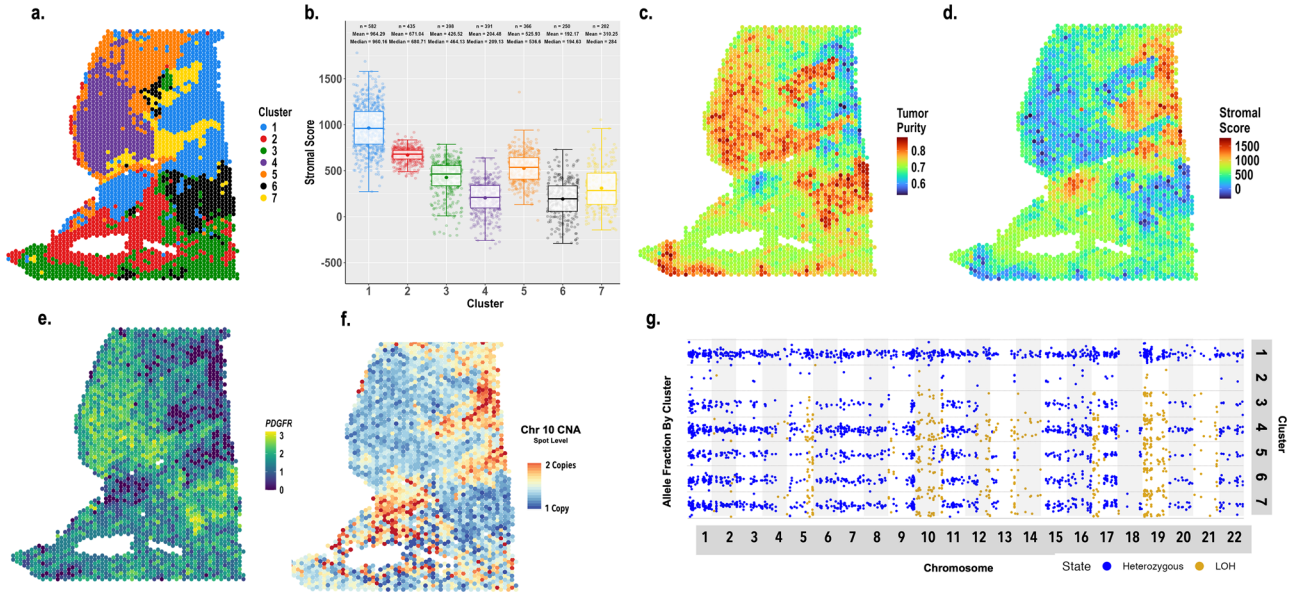

Fig. 4 | Spatial distribution of stromal and tumor in Diffuse Midline Glioma A1. **a** Graph clustering assignment spatial map of Diffuse Midline Glioma A1, an *IDH* wild-type, *EGFR* negative tumor. **b** Boxplots of ESTIMATE stromal scores for each cluster. The minima and maxima are noted by the whiskers for each boxplot. The 25ᵗh percentile and 75th percentile are shown as the bounds of each box. A line across each box represents the median value. The minima is calculated as Q1-1.5*(interquartile range), and the maxima is calculated as Q3 + 1.5*(interquartile range). The mean for each cluster is shown as a point on each boxplot. Mean, median, and number of observations are displayed above each plot. All data points are shown behind each boxplot. **c** ESTIMATE tumor purity spatial overlay. **d** ESTIMATE stromal score spatial overlay. **e** *PDGFRA* log normalized counts spatial overlay. **f** Copies of chromosome 10 per spot, whereby the average of normalized gene expression for all genes above the average of 0.1 counts, where tumor regions are expected to have chromosome 10 deleted. **g** tLOH output of allele fractions for SNPs with total read counts greater than 20. Points are highlighted based on state determination, where blue is heterozygous, and gold is LOH.

Validation of *EGFR*/p53 genomic spatial heterogeneity in a separate cohort

The results from the discovery cohort revealed how ecDNA contributes to spatial heterogeneity in glioblastoma, allowing cancer cells to rapidly reprogram their genome at the DNA level in response to changes in the microenvironment. This spatial heterogeneity was also associated with additional genomic alterations, including the LOH on chromosome 17 in the *TP53*-mutated Glioblastoma A1. To further explore these findings, we performed exome sequencing on tumor DNA from an additional set of six *IDH* wild-type glioblastoma samples. Of these, six exhibited focal ecDNA-driven DM amplification of *EGFR*, forming the validation cohort (Glioblastoma B1 through B6), as shown in Fig. 6 and Supplementary Fig. 9. Samples were spatially profiled using the Visium CytAssist Spatial Gene Expression assay for formalin-fixed paraffin-embedded (FFPE) tissue. This assay is an adaptation of the Visium platform used earlier, enabling the analysis of the more broadly available FFPE tissue. One relevant change is that this assay version uses a probe-based method to capture RNA fragments. Consequentially, using probe-based capture and ligation for LOH analysis through expressed SNPs is no longer feasible in the FFPE-derived RNA. Instead, we inferred copy number variations by correlating gene expression patterns with the ecDNA DMs identified through exome sequencing.

Co-amplification of *MDMX*

Two of the six *EGFR*-amplified samples showed a second independent DM harboring negative regulators of the p53 pathway, specifically *MDM2* or *MDM4* (Fig. 6a), based on bulk tumor DNA coverage and breakpoint analysis. For example, in the case of Glioblastoma B7, which presented with an *EGFR* DM from chromosome 7p11, a second DM was found on chromosome 1q32.1 (from 202 to 205 Mb), notably containing the gene *MDM4*. Coverage analysis indicated that this sample carried approximately 18+ copies of *EGFR* and 8 copies of *MDM4*. In a similar case, Glioblastoma B11 revealed a second 230 kb ecDNA DM on chromosome 12q15, which housed the *MDM2* gene. Bulk DNA analysis suggested that Glioblastoma B11 contained around 15 copies of *MDM2* and over 30 copies of *EGFR*.

The amplification of *MDM2* and *MDM4* (collectively referred to as *MDMX* mutations) is a well-established mechanism by which glioblastomas inactivate p53. *MDM2* and *MDM4* are known to negatively regulate the *TP53* tumor suppressor by either promoting its degradation or inhibiting its transcriptional activity, thus leading to unchecked cell proliferation and survival. In the case of Glioblastoma A1, the loss of *TP53* was observed, which is mirrored in the validation set, where *MDMX* amplification plays a similar role by functionally inactivating the p53 pathway downstream. This co-amplification of *EGFR* with either *MDM2* or *MDM4* is common in glioblastomas and pivotal to tumor progression, and it disrupts the normal cell-cycle checkpoints enforced by p53.

Spatial gene expression heterogeneity in *MDMX*/*EGFR*

Within both Glioblastoma B7 and Glioblastoma B11, we observe distinct tumor regions characterized by high *EGFR* expression (Fig. 6). Within a subset of these *EGFR*-overexpressing regions, elevated *MDM4* or *MDM2* expression also co-occurs. This observation is consistent with findings from our earlier discovery cohort, specifically in Glioblastoma A1, where *EGFR* overexpression was associated with loss of heterozygosity on chromosome 17, harboring the mutated *TP53*.

In areas where both *EGFR* and *MDMX* are highly expressed, we observe a marked enrichment of proliferative gene signatures. This suggests that dual overexpression of these two oncogenes may enhance tumor cell proliferation, potentially through pathways involved in DNA repair, apoptosis resistance, or unchecked cell cycle progression. The increased proliferative activity in these regions aligns with previous studies implicating *EGFR* and *MDMX* in promoting aggressive tumor phenotypes, underscoring their role as key drivers of glioblastoma growth and therapy resistance. In

https://doi.org/10.1038/s41467-025-59805-z

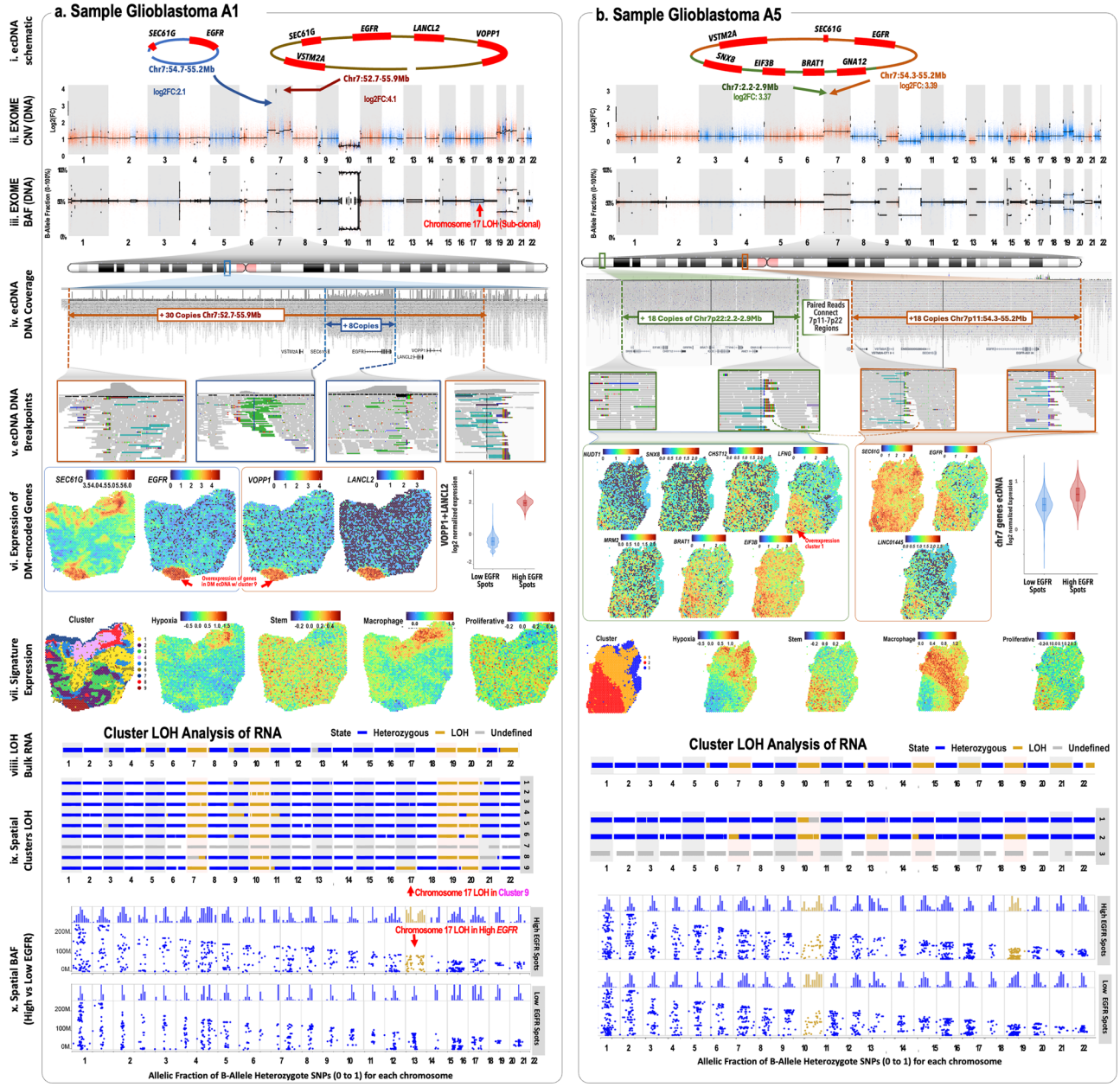

Fig. 5 | Subclonal genomic heterogeneity from extrachromosomal DNA double-minutes. Spatial transcriptomic analysis of **a** Glioblastoma A1 (left panels) and **b** Glioblastoma A5 (right panels) reveals spatially distinct subclonal genomic heterogeneity, demonstrating loss of heterozygosity in regions that also show overexpression of ecDNA-amplified genes, including *EGFR*. As discussed in the text, Glioblastoma A1 showed LOH of chromosome 17 within high *EGFR*-expressing cluster 9. For each vertical panel: **i** Schematic of EGFR+ amplified ecDNA fragments based on exome copy number and breakpoint analysis. Glioblastoma A2 shows two different EGFR+ ecDNA variants. **ii**–**iii** Log2 fold change and B-allele frequency plots from exome sequencing CNV and LOH analysis. **iv**–**v** Exome analysis coverage of ecDNA, indicating breakpoints and mapped regions. **vi** SCT-normalized spatial expression of genes within ecDNA regions. Violin plot shows the expression of normalized coverage of genes encompassing DM ecDNA (excluding *EGFR*) for high *EGFR* and low *EGFR* spots (low *EGFR* < = 5 counts). **vii**. Spatially mapped average expression for gene expression marker of hypoxia, proliferative cells, macrophages, and cancer stem cells. **viii**. tLOH analysis showing spatially distinct clusters with LOH. **ix**–**x** Spot-level B-allele frequency analysis indicates clonal LOH of chromosome 17 in recurrent Glioblastoma A1 and LOH on chromosome 19 in areas with high *EGFR* expression (high *EGFR* > 8 counts). Bars and points are color-coded by state: blue for heterozygous, gold for LOH, and gray for undefined.

contrast, in regions where *MDMX* expression is moderate or lower, *EGFR* expression remains high and is additionally associated with a vascular hypoxia signature (Fig. 6i, iv).This shift in gene expression may reflect the tumor's adaptation to hypoxic microenvironments, where *EGFR*-driven signaling is uncoupled from *MDMX*-mediated pathways. In these hypoxic regions, the tumor likely relies on angiogenic and metabolic reprogramming to survive, with lower *MDMX* levels perhaps indicating reduced apoptotic inhibition.

Genomic reprogramming through ecDNA and chromosomal alterations drives spatial heterogeneity

The dynamic nature of ecDNA DMs may enable cancer cells to adaptively reprogram their genomes, driving tumor progression in response to changes in their microenvironment. Analysis of *EGFR*-amplified, *IDH* wild-type glioblastomas revealed four instances where spatial gene expression heterogeneity suggests that the underlying mechanisms involve genomic alterations either through LOH, such as in Glioblastoma A1, or through changes in the copy number of key

https://doi.org/10.1038/s41467-025-59805-z

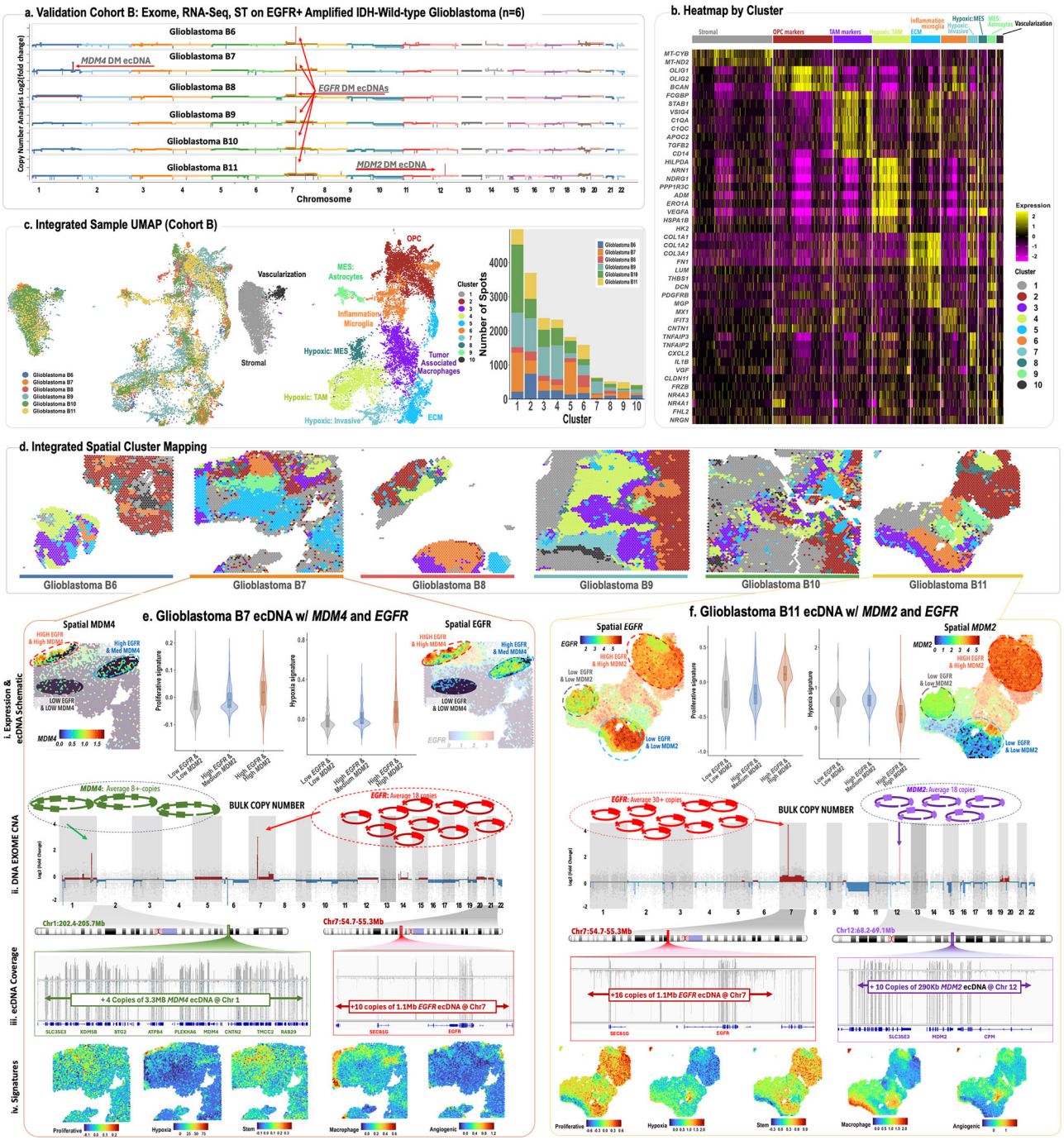

Fig. 6 | Validation cohort B examining transcriptomic and genomic subclonal heterogeneity in *EGFR*+ glioblastoma. Exome sequencing, bulk RNA sequencing, and FFPE ST were used in the validation set, following a similar analysis to cohort A. **a** Copy number analysis through exome sequencing, showing the location of focal amplifications. **b** Expression heatmap of top marker genes per cluster. **c** UMAP plot of integrated dataset, color-coded by sample, cluster, and the distribution of samples within integrated clusters. **d** Spatial maps of integrated cluster assignments for each sample. While all six samples contained amplified *EGFR* within a DM ecDNA, two of the six samples also had amplification of negative regulators of *TP53*. **e** Glioblastoma B7 contains independent DMs of *MDM4* and *EGFR*. **f** Glioblastoma B11, showing independent DMs with *MDM2* and *EGFR*. Evidence for DMs based on copy number analysis and corresponding expression data are shown in both panels: **i** Spatial expression of *MDM4*/*MDM2* and *EGFR*, alone with violin plots of hypoxia and proliferative signatures, **ii** Detailed copy number analysis, showing focal amplification zooms, **iii** Coverage at locus, **iv** Expression profiles for key pathways highlight that regions expressing genes within the DM regions in both samples also align with areas of high proliferative and hypoxic gene signatures.

oncogenes located within DMs. These findings highlight a recurrent pattern of co-amplification between *EGFR* and inactivating p53 genomic mutations, linking recurrent structural mutations to the broader context of genomic instability that is a hallmark of glioblastoma. In conclusion, identifying subclonal genomic alterations, especially those driven by ecDNA, provides key insights into glioma biology. While loss of heterozygosity exemplifies one such alteration, ecDNA amplifications, particularly involving potent oncogenes like *EGFR* and *PDGFRA*, serve as fundamental drivers of the genetic instability and heterogeneity underpinning these tumors' aggressive nature.

Discussion

Gliomas are characterized by their complexity and spatial heterogeneity, with diverse populations of glial cells, tumor stem cells, and subtype-defining somatic alterations distributed across different regions. Mapping the spatial organization of genomic and transcriptomic alterations is crucial for understanding the molecular processes driving tumor pathogenesis in these highly heterogeneous tumors. In this study, we characterize the spatial transcriptomic heterogeneity of various gliomas, including glioblastomas, oligodendrogliomas, astrocytomas, and diffuse midline gliomas, using ST alongside germline blood, tumor DNA, and tumor RNA sequencing. Through this integrative approach, we identify spatially distinct tumor subclones driven by somatic aneuploidy using loss of heterozygosity analysis, as well as variation in extrachromosomal DNA double minutes harboring key oncogenes like *EGFR* and *TP53* pathway regulators, *MDM2* and *MDM4*, in glioblastoma. The DNA-level spatial heterogeneity within these subclones provides a mechanism for genomic plasticity, driven by recurrent and spatially distinct *EGFR* amplifications of p53 inactivating mutations, thereby enhancing the tumor's ability to adapt to selective pressures.

Clustering across gliomas

We have mapped common motifs through integrated clustering across the major types of high-grade glioblastomas, oligodendrogliomas, astrocytomas, and diffuse midline gliomas. Our analysis of 11 samples revealed gene expression patterns that vary according to the type of glioma. For instance, we noted areas within tumors where rapid growth may lead to oxygen deprivation, as evidenced by adjacent clusters showing angiogenesis (marked green in Fig. 2d) and hypoxia (marked red in Fig. 2). In some gliomas, *PDGFRA* expression, associated with oligodendrocytes and new blood vessel formation, was opposite to *HIF1A*, a hypoxia indicator. Additional antibody-based methods will be important to show whether this is a generalizable feature.

Specifically, in glioblastomas, astrocytomas, and diffuse midline gliomas, we identified regions where glial indicators like *VEGFA* and *VIM* are close to hypoxic areas with elevated levels of *NDRG1* and *HSPA1B*. In contrast, oligodendrogliomas showed increased levels of *SMOC1*, *OLIG1*, *MAG*, and *MOG*, linked to oligodendrocytes and their precursors. We do see evidence that the cluster of the current 55-micron spots may be influenced by mixtures of multiple cells; in particular, the mixed lineage cluster showed fewer marker genes and less distinct, with expression of oligodendrocytes, oligodendrocyte precursor cell markers, and in a small set of spots, neural precursor cell markers. Each tumor exhibits unique tissue complexity, and integrated clustering may obscure certain features. For instance, Glioblastoma A1 displays distinct tissue complexity through graph clustering. Within this tumor section, several disease-relevant genes, including *VEGFA*, *HIF1A*, *PDGFRA*, *NDRG1*, *DDIT4*, and *CD44*, demonstrate varied expression distributions across tissue sections distinctly different from integrated clustering (Supplementary Fig. 6).

Tracking chromosome gains and losses with loss of heterozygosity

Observing tumor subclones with distinct mutational profiles represents a significant advancement in understanding how ST can provide insights into tumor pathogenesis. We utilized germline sequencing to identify LOH within the fresh-frozen ST samples. Other groups have explored these methods on single-cell and spatial transcriptomics, and they are likely essential to understanding tumor evolution and pathogenesis. It is important to highlight the emergence of multiple tools and methods in these areas[29–34]. While we considered varied tools and packages in our study, each displayed its own unique set of biases and strengths, which is anticipated given the rise of new experimental techniques and innovative analytical strategies. Our results highlight the efficacy of LOH analysis in deducing tumor content and even in detecting subclonal tumor genomic heterogeneity, which is directly related to tumor pathogenesis.

Nevertheless, discussing the challenges that impact the interpretation of LOH analysis with ST data is pertinent, including low read coverage, multiple cells within a capture spot, variability in sequencing saturation, and gene capture variability. Across our dataset, the average number of genes detected was around twenty thousand. However, the total reads fluctuated between 130 million and 390 million. A significantly higher read coverage is essential to reporting LOH in conventional copy number analysis. This requisite coverage becomes even higher for bulk tumor-only analysis, demanding at least 200x coverage over the sites of interest to detect heterozygous SNPs and to see allelic subtle shifts within tissues.

Leveraging bulk tumor sequencing to characterize genomic spatial heterogeneity linked to ecDNA DMs

The main application of inferring genomic copy numbers in tumors has been to differentiate between normal and tumor tissues, with fewer instances of its use in identifying true genomic subclones, as demonstrated in recent prostate tumor studies[35]. A likely reason for the lack of spatial subclonal examples is partly because genomic alterations that drive pathogenesis would be rare events under clonal expansion models and less likely to be observed in a 6.5 mm by 6.5 mm grid for a single tissue section. Tumor progression often involves a recurring series of genomic gains and losses, contributing to subclonal heterogeneity and localized regions of tumor advancement. These events can be frequent at the local level. In particular, ecDNA DMs vary in copy number and frequently harbor genes that drive tumor progression. This often includes amplified *EGFR* in gliomas, commonly found within DMs, which enhances proliferative signaling and accelerates tumor growth[22]. Notably, in our analysis, Glioblastoma A1, Glioblastoma A2, and Glioblastoma A5 exhibited high levels of *EGFR* amplification, while Diffuse Midline Glioma A1 and Glioblastoma A4 showed significant amplification of *PDGFRA*.

Spatial DNA genomic heterogeneity along the *EGFR*/p53 axis

The importance of spatially mapping genomic DNA alterations driving tumor progression is exemplified in Glioblastoma A1, which is positive for *EGFR*vIII, a transcript variant with deleted exons 2–7, known to promote cell proliferation, aggressiveness, and tumor growth through constitutively active signaling[36]. ST analysis reveals a spatially distinct cluster with elevated *EGFR* expression and a complete loss of heterozygosity on chromosome 17 containing an initial *TP53*(R252C) mutation. First, clinical reports show *EGFR* amplification is subclonal based on copy number and allelic analysis. More significantly, bulk sequencing reveals a modest LOH on chromosome 17, with the phasing and parental copy loss mirroring the spatial transcriptomics subclone. This matching allelic loss pattern between the hypoxic subpopulation and overall tumor suggests the genomic changes in the *EGFR* 17p LOH subclone are prerequisites for forming hyper-nucleated hotspots exhibiting necrosis and cell death.

Sequencing a second set of six *IDH* wild-type, *EGFR* amplified glioblastoma samples showed spatial heterogeneity of ecDNA DMs harboring amplification of key negative regulators of the p53 pathway in two samples: *MDM4* in Glioblastoma B7 and *MDM2* in Glioblastoma B11. These ecDNA DMs highlight the genomic plasticity, allowing distinct regions with amplified *EGFR* and other areas where both *EGFR* and *MDM2* are overexpressed. In regions where only *EGFR* is amplified, we primarily observed an enrichment of hypoxic gene signatures. These hypoxic zones may represent adaptations to a poorly oxygenated microenvironment, contributing to glioblastoma's vascular remodeling and neovascularization characteristics.

In contrast, regions where both *EGFR* is overexpressed and p53 is inactivated, either through *TP53* loss or *MDM2*/*MDM4* amplification, display a different profile. Here, we find that a pronounced

proliferative gene signature with reduced hypoxic markers, indicating that these areas may be more proliferative and less reliant on hypoxia-driven pathways. This suggests that concurrently overexpression of these two oncogenes may enhance tumor cell proliferation, potentially through pathways involved in DNA repair, apoptosis resistance, or unchecked cell cycle progression. The observe hypoxia signature in neighboring zones could contribute to the vascular remodeling and neovascularization often seen in glioblastoma[9,13,37]. Taken together, this spatial heterogeneity in *MDMX* and *EGFR* expression underscores the dynamic interplay between proliferative and survival pathways in distinct tumor microenvironments, each presenting unique therapeutic challenges. Moreover, the additional loss of p53-mediated cell cycle regulation and genomic instability from LOH of *TP53* or *MDM2*/*MDM4* amplifications likely drive multinucleation, polyploidy, and chromosomal instability. This genomic instability and *EGFR* expression may create further treatment-resistant subclones.

Further experimental design considerations

Our results suggest several important considerations for future design. Experimentally, in-depth sequencing and multiple sections could yield more reliable results for future studies. As spatial technology progresses, we can expect better read depth and resolution. Individual samples might need specific modifications due to the inherent variability of this budding methodology. Generally, sequencing ST sections more comprehensively and ensuring multiple sections per tumor would enhance the general applicability of the results.

Taken together and overall, by integrating spatial transcriptomics and exome sequencing, we identified spatially distinct subclonal events, such as ecDNA DMs and LOH, and characterized these alterations in the context of glioblastoma progression. The replication of these findings across multiple samples highlights the importance of ecDNA and p53 pathway regulators in common glioblastoma pathways, offering deeper insights into potential therapeutic targets, particularly in *EGFR* positive *IDH* wild-type glioblastomas.

Methods

Glioma tumor tissue collection

Eleven fresh frozen, surgically derived glioma biopsy specimens of varied grading and pathology were analyzed in cohort A. Six *IDH* wild-type, *EGFR* amplified, surgically derived glioblastoma FFPE biopsy specimens were analyzed in cohort B. Cohort A includes glioma tissues from male and female individuals between the ages of 25 and 63 at the time of tissue collection. Cohort B includes glioblastoma tissues from male and female individuals between the ages of 30 and 76. Limited individual-level sample information including age and sex is available in Supplementary Data 7. Individual sex was self-reported. Study design was determined based on glioma pathology and molecular phenotype. De-identified human samples were obtained from the University of Southern California Norris Comprehensive Cancer Center Translational Pathology Core Facility, and written informed consent for analysis and publication of results was collected from every participant. Approval was obtained for this study by the University of Southern California Biomedical IRB (IRB protocol HS #11-00385).

Exome library preparation and sequencing for cohort A

DNA was isolated from adjacent fresh frozen tissue of each glioma tumor block, and germline DNA was isolated from whole blood. For cohort A, the Qiagen blood protocol (Qiagen, #69504) was used with minor adjustments to account for 100 mL initial input volume. Protease was increased 1.5-fold, and protease incubation time was increased from 10 min to 60 min at 56 °C. Ethanol precipitation was increased from 1 min to 30 min. The number of samples simultaneously centrifuged was decreased to 4. Three washes were added to the end of the protocol to remove residual heparin. The final elution was reduced from 200 mL to 50 mL using pre-heated (56 °C) H_2O. The resulting isolates from tumors and blood were analyzed by standard Illumina whole exome sequencing using Agilent OneSeq V6 + UTR enrichment probes (Agilent Technologies, Inc., #5190-8888). Library construction was performed with Agilent SureSelect XT-LI dual indices and enzymatic fragmentation (Agilent Technologies, Inc., #G9916A and #5191-4080). Sequencing was performed on an Illumina NovaSeq 6000 system.

RNA library preparation and sequencing for cohort A

RNA was isolated from adjacent fresh frozen tissue of each glioma spatial sample tumor block. The NEBNext Ultra II directional RNA kit was used, with rRNA depletion and dual indices (New England Biolabs, #E7760S). Sequencing was performed on an Illumina NovaSeq 6000 System.

Fresh frozen Visium spatial transcriptomic analysis for cohort A

The 10X Genomics Visium Spatial Gene Expression method (10X Genomics, Inc., #1000187) consists of cryosectioning, tissue fixation to custom gene expression slides, hematoxylin and eosin (H&E) staining, high-resolution imaging, tissue permeabilization, reverse transcription, second strand synthesis, cDNA denaturation, cDNA amplification, and sequencing library construction. At the end of this process, each sequencing read contains a 16 base pair barcode sequence and 12 base pair unique molecular identifier (UMI) to relate the data to coordinates on each spatial grid. Sample libraries were pooled and sequenced on an Illumina NovaSeq 6000 System. The specific protocol used was CG000239 RevC.

Glioma tumor blocks were prepared for the cryosectioning step by the USC Norris Comprehensive Cancer Center Translational Pathology Core Facility using the 10X Genomics Tissue Preparation Guide. Tissues were sectioned at 10 µm thickness and adhered to one of four available 6.5 × 6.5 mm grids on the gene expression slides. Following H&E staining of the tissue, high-resolution imaging was performed was performed by the USC Stem Cell Optical Imaging Facility. Tissue permeabilization was performed for 12 min, which was determined to be optimal by the 10X Genomics Tissue Optimization Procedure. Spatial data was bioinformatically analyzed using the 10X Genomics spaceranger pipeline v1.1.0 (https://www.10xgenomics.com/support/software/space-ranger/latest). The analysis consisted of read alignment to GRCh38 with STAR, unique molecular identifier counting, principal component analysis, construction of a k-nearest neighbors graph, and gene expression clustering (graph and k-means).

Exome library preparation and sequencing for cohort B

DNA was isolated from adjacent FFPE tissue and germline DNA was isolated from whole blood for each glioblastoma sample. Samples were prepared using standard protocols for the Twist Bioscience for Illumina Exome 2.0 Plus Panel (Illumina, Inc., #20076914), Illumina DNA Prep with Enrichment, (S) Tagmentation (Illumina, Inc., #20025523), and IDT for Illumina DNA/RNA UD indexes Set A, Tagmentation (Illumina, Inc., #20027213).

RNA library preparation and sequencing for cohort B

RNA was isolated from adjacent FFPE tissue of each glioblastoma spatial sample tumor block. Samples were prepared with the Takara SMARTer Stranded Total RNA-Seq Kit v2 – Pico Input Mammalian with HT for Illumina v2 (Takara Bio USA, Inc., #634412) protocol.

FFPE Visium CytAssist spatial transcriptomic analysis for cohort B

Samples in cohort B were prepared according to the 10X Genomics Visium CytAssist (10X Genomics, Inc., 10000520) protocols CG000520 RevB and CG000495 RevC. Sample libraries were sequenced on an Illumina NextSeq 2000 P3 System. FFPE glioblastoma

tissue blocks were selected and sectioned at 5 µm thickness, then placed individually on tissue slides. Slides were stained with hematoxylin and eosin (H&E) to visualize tissue morphology, followed by differentiation in 1% acid alcohol, bluing in 1x PBS, and dehydration in graded ethanol and xylene. Slides were loaded into the CytAssist, and the transfer process was conducted following the manufacturer's protocol to ensure precise capture of spatially resolved mRNA molecules. After transfer, the reverse transcription step was performed to synthesize cDNA from captured mRNA on the tissue slide. Tissue sections were removed, and spatially barcoded cDNA libraries were constructed and sequenced following the 10X Genomics protocol. Spatial data was bioinformatically analyzed with the 10X Genomics spaceranger tool v2.0.1, and reads were aligned to the GRCh38 reference. The Visium Human Transcriptome Probe Set version was v2.0.

Exome sequencing bioinformatic analysis for cohort A and cohort B

After sequencing, FASTQs were aligned to GRCh38 using bwa-mem v0.7.17[38]. GATK v4.0.10.1 IndelRealigner minimized mismatches across local alignments, and GCBias determined coverage bias[39]. GATK CollectMultiMetrics and samtools v1.9 stats were used to output sequencing statistics[40]. The tumor and germline BAMs were joint analyzed with GATK HaplotypeCaller and SnpEff v4.3[41]. The dbSNP v146 hg38 VCF was input as a reference for annotation[42]. Copy number analysis was performed with the tool tCoNuT v1.0 (https://github.com/tgen/tCoNuT) and GATK Somatic Copy Number Variant Discovery v4.2.6.1[39].

Bulk RNA sequencing bioinformatic analysis for cohort A and cohort B

After sequencing, FASTQs were aligned to GRCh38 using STAR v2.6.1d[43]. Aligned BAMs were then processed with GATK MarkDuplicates v4.0.10.1, then sorted and indexed with samtools. Summary statistics were obtained using samtools stats and Picard RNA metrics (GATK). Salmon v0.11.3, FeatureCounts v1.6.3, and HTSeqCounts 0.6.0 were used to output gene and transcript level counts from the aligned reads[44–46].

Spatial transcriptomics data integration for glioma cohort A

ST data was normalized with the R package Seurat's (v4.3.0) SCTransform normalization, after which the effects of mitochondrial percentage were regressed out, and all samples were integrated by reciprocal PCA[47,48]. A principal component analysis with 50 principal components was then performed on the integrated dataset. The FindNeighbors Seurat function was applied, followed by FindClusters with a resolution of 0.2. This resolution was determined to be optimal after iteration of a wide range of clustering resolutions and examination of cluster markers and barcode distributions. The RunUMAP Seurat function was applied to obtain a UMAP projection, plotted separately by sample and cluster (Fig. 2a,b). Markers for the per cluster heatmap in Fig. 2d were calculated using the FindAllMarkers function.

Integrated glioma cohort A feature analysis

To generate a heatmap of relative expression of the transcriptional programs published by Ren et al. for Fig. 1f, we imported their published gene sets in R[8]. We applied the Seurat AddModuleScore function for each set and plotted scores using ComplexHeatmap v2.14.0[49]. The same method was applied for the heatmap of relative expression of the spatial-specific transcriptional programs published by Ravi et al. in Fig. 2g[9].

Spatial BAM separation and allele count calculation

A python pipeline was implemented to prepare cohort A spatial BAM files for Bayes factor calculation and hidden Markov analysis. The objective of this pipeline is to obtain per-cluster coverage for reference and alternative alleles at heterozygous SNP positions. Using this method, all spaceranger output BAMs from the glioma cohort A spatial data were split into individual per cluster BAM files. Primary alignments were extracted with samtools v1.9. Candidate SNP positions detailed below were used as a reference for coverage calculation. A VCF formatted file was output for each sample, with graph-based clusters listed as individual samples. Positions with total coverage above 2000 unique molecular identifier counts were removed due to potential errors that could occur with integration during Bayes calculations. Positions with total coverage less than 5 total reads are also removed to reduce false positive results. SNPs within the human leukocyte antigen region (chr6:28,510,120-33,500,500) were also omitted. The packages utilized in this method include NumPy v1.19.5, Pandas v1.1.5, and Pysam v0.15.4 (https://github.com/pysam-developers/pysam)[50,51]. The diagrams describing this process in Fig. 3a, e were generated using Microsoft PowerPoint.

Selection of candidate heterozygous SNP positions

A filtered database VCF was generated to identify likely heterozygous SNPs in the population for reference in the BAM separation step. The dbSNP common database (human_9606_b151_GRCh38p7) VCF was obtained from the NCBI FTP download site. Next, the file was annotated for canonical transcripts with SnpEff, followed by filtering for transcripts in the untranslated regions (UTR). SNP positions with minor allele frequency less than 0.10. SNP positions where the allele fractions were greater than 0.10 or less than 0.90 were kept for further analysis, ensuring that the remaining SNPs were likely heterozygous. To validate the results of this approach on the glioma cohort A spatial data, reference SNP positions were also separately determined from the companion bulk exome sequencing. HaplotypeCaller VCFs containing both tumor and germline samples from joint calling were used for this validation. The dbSNP reference-dependent and exome sample-specific VCFs were annotated for canonical transcripts by SnpEff and filtered for variants containing the terms UTR, missense, nor nonsynonymous. Positions were further filtered to exclude multiallelic calls, indels, and positions without a dbSNP rsID identifier. The following methods were tested on tumor-only and bulk exome reference approaches. The workflow of BAM separation, allele count calculation, and SNP selection is available as a set of python scripts.

Bayesian model

The Bioconductor package tLOH, or transcriptomicsLOH, is a workflow of R functions which leverages Bayes factors and a hidden Markov model for determining regions of LOH in ST data. The base input is a VCF with each cluster listed as a sample. Data is imported using the tLOHImportData function, which reads the VCF into a data frame. Next, the function tLOHCalc computes Bayes factor likelihood ratios of two models at every SNP position. The allele count data is assumed to have a binomial distribution, where each SNP position is an independent event. The two possible outcomes of the binomial are a normal or LOH state. Each independent event results in a success or failure in n trials. Measurement of success would be reads supporting the reference allele, and failure would be reads supporting the alternative allele. The number of trials n is the total reads at a given SNP position, and x is the number of reads aligning to the reference allele. In our equation, Model 1 (M1) represents a loss of heterozygosity event, and Model 2 (M2) represents a heterozygous state. D represents the data which informs each model. Bayes factor K likelihood ratios are calculated at each SNP to support evidence for either model.

$$K = \frac{P(M1|D)P(M2)}{P(D)P(M1)} \quad (1)$$

Model 1 is a loss of heterozygosity event where the prior probability is determined through a modified beta distribution where $0 \le \theta \le 1$,

α=500, and β=500.

$$P(\theta) = \frac{abs(0.5-\theta)^{\alpha-1}(1-abs(0.5-\theta))^{\beta-1}}{B(\alpha,\beta)} \quad (2)$$

Model 2 is a heterozygous event where the prior probability is determined through a beta distribution where $0 \le \theta \le 1$, $\alpha = 1.25$, and $\beta = 1.25$.

$$P(\theta) = \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha,\beta)} \quad (3)$$

The expected probability of a model given the data is the product of the binomial distribution and the beta distribution. The formula is multiplied by an error factor $e = 0.01$ and subtracted from one. The formula below is the probability of Model 2 given the data.

$$P(D) = \int_0^1 \theta^x(1-\theta)^{n-x}(1-e)\frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha,\beta)}d\theta \quad (4)$$

For both models, the marginal likelihoods are obtained by multiplying the likelihood with the prior distribution and integrating it to obtain the area under the curve. The beta distribution is used as the prior distribution of the binomial. Bayes factors at each genomic position are calculated by the division of $P(M_{loh+}|D)$ and $P(M_{loh-}|D)$.

$$K = \frac{P(M_{loh+}|D)}{P(M_{loh-}|D)} \quad (5)$$

After calculation of Bayes factors, values are log transformed. The threshold for likelihood of a model is determined by reference to Appendix B of Theory of Probability by Harold Jeffreys[52]. Log 10 K values greater than or equal to 0.5 represent substantial evidence for Model 1. Log 10 K values less than or equal to −0.5 represent substantial evidence for Model 2. Values between 0.5 and −0.5 are nondeterminate.

A hidden Markov model was implemented to classify regions as normal or loss of heterozygosity with the likelihood ratios as prior information. There are two states in the HMM. The number of states were determined by the nature of the model of either LOH (state 1) or HET (state 2), where the likelihood of LOH vs heterozygous event is computed at a SNP position.

The HMM analyzed ordered quantile normalized Bayes factor values in each chromosome in each cluster. Bayes factors are calculated from allele counts, followed by an ordered quantile normalization to ensure a normal distribution in each chromosome. A 2-state HMM then segments each region, outputting state assignments. Cumulative metrics across each segment are calculated, and a state label is applied. We did not provide individual emission parameters for the HMM, they were calculated for each iteration by the R package depmixS4.

The tool depmixS4 was used for HMM implementation[53]. In our method, the allele count and Bayes factor data frames are split into per-chromosome and per-cluster sets for analysis. Ordered quantile normalization is applied to each set, transforming the likelihood ratios to a normal distribution[54]. The R package bestNormalize orderNorm function is applied (equation shown below), where Φ is the standard normal cumulative distribution function, and x is each observation[55].

$$g(x) = \Phi^{-1}\frac{rank(x)-0.5}{length(x)} \quad (6)$$

The default settings of the initial state is set per chromosome manner. The two-column table describing these parameters is available at [https://github.com/USCDTG/tLOH/blob/main/data/initialStartProbabilities.rda]. The transition parameters are defined as a matrix of the following values: c(0.999,0.001), c(0.001,0.999). The tLOH function to run the HMM processing is hiddenMarkovAnalysis, applied to output from the tLOHCalc function.

Baum-Welch was used to determine the initial start probabilities, and training was conducted on bulk RNA data. Several iterations of training were run to determine parameters, including using a convergence threshold set to a defined threshold. Once HMM was trained for all gliomas, we examined the stability of threshold and lengths of segments to assess agreement with segments from independent bulk copy numbers. In practice, training took several hours and typically yielded multiple segment options. The emission probability matrices were set as default, though our method provides an option to provide custom parameters. The main functions used from the depmixS4 package are depmix, fit, em.control, and posterior. The decoding algorithm specified for fit and posterior is Viterbi. Cumulative metrics are evaluated across defined HMM segments to obtain final state assignment labels.

After running the HMM on the per-cluster, per-chromosome data, the state labels and probabilities are merged into a data frame containing all clusters and chromosomes. The final output is a sample.csv with per-segment metrics, individual SNP Bayes factors, and state assignments. Plots representing output from our method were generated using ggplot2 v3.4.2[56]. The diagram describing the HMM in Supplementary Fig. 8 was generated with standard Microsoft PowerPoint shapes.

Validation of loss of heterozygosity

The bulk exome copy number variation data was analyzed using two tools: tCoNuT v1.0 and GATK Somatic Copy Number Variant Discovery v4.2.6.1. To compare the spatial method to the bulk exome data, we first analyzed complete (not split) spatial BAMs with our LOH identification method in R. After state determination, we transformed the output data to a.bed file format with a measurement of every million bases for each chromosome. We labeled bases with the state determinations if they fell between the start and end positions of the HMM segments. Next, file output from the GATK tool was read into R and split into 1 Mb segments. We then assigned a state label of LOH if the exome segment mean value was below a predefined level. Positions across the spatial and exome sequencing results were compared, and regions with no overlap were omitted. Metrics of specificity, sensitivity, positive predictive value, false discovery rate, and percent reported were calculated (Supplementary Data 4).

Visualization of SNP density

The SNP density plot in Fig. 3c was generated using the R packages KaryoploteR v1.30.0, regioneR v1.36.0, and GenomicRanges v1.56.2[57–59]. We filtered for unique SNPs across the LOH analysis dataset and converted the list of genomic positions to a Granges object. The plotKaryotype and kpPlotDensity functions were applied with customized visualization parameters.

Minimal detectable event size

A central objective in our method development was to determine the limits of detection of the core algorithm and the minimum data necessary to define a valid result. Key metrics used in state determination were the median segment K value, the sequential sum of log10(K), and the segment mode peak. We required at least an interval length greater than 1000 bases for a valid segment. HMM segment lengths across the dataset range from 1000 to 248,000,000 bases. In the case of Fig. 5, several regions were listed as 'undefined' due to failure at the HMM step, but the allele fractions and segment values

strongly suggested a complete heterozygous state. An improvement to our tool would be the implementation of an additional check that first identifies the likelihood of two states and allows segmentation is there is supporting evidence. Alternatively, we could allow for more algorithm runs for convergence or adjust the starting transition probabilities in response.

Tumor purity analysis

'Estimation of STromal and Immune cells in Malignant Tumors using Expression Data' or ESTIMATE v1.0.13, was run on log normalized spatial unique molecular identifier counts to investigate tumor purity and stromal scores across samples[26]. ESTIMATE plots for all samples in Supplementary Fig. 7 were generated using Seurat, ggplot2, and the viridis color map package v0.5.6[60].

Spatial transcriptomics data integration for cohort B

The FFPE ST dataset was normalized with the R package Seurat SCTransform normalization. Samples were integrated using the RunHarmony method v1.2.1[48,61]. A principal component analysis with 30 principal components was then performed on the integrated dataset. The FindNeighbors Seurat function was applied, followed by FindClusters with a resolution of 0.19. Markers were calculated using the FindAllMarkers function[61]. Gene signatures described in Fig. 6b and Supplementary Data 5 were based on prior literature[8,11,13,15].

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The processed spatial transcriptomics sequencing data generated in this study is available in the NCBI Gene Expression Omnibus repository under accession code GSE242352. The raw sequencing data generated in this study is available in the European Genome-Phenome Archive (EGA) under restricted access at [https://ega-archive.org/datasets/EGAD50000001394]. Raw sequencing data is stored under restricted access to protect data privacy. Access to this data may be requested through the EGA website by submitting a form to the Data Access Committee (DAC) EGAC50000000579. Restrictions for access to the data are outlined by the EGA DAC Policy EGAP50000000519. The DAC policy information states terms and conditions which include preservation and protection of sequencing data confidentiality, appropriate use of data in only non-commercial and academic purposes, acknowledgment of data use, terms for data management upon completion of research, and compliance with applicable laws, regulations, and ethical guidelines. The expected response timeframe for access requested to the DAC will be 1 to 2 weeks. Data will be available for 1 year once access has been granted. Source data are provided with this paper.

Code availability

Analysis code and documentation for the results described in this work are available for download at [https://github.com/USCDTG/paperAnalysis_LOH]. The transcriptomicsLOH R package is available for download at [https://github.com/USCDTG/tLOH] under the MIT license. Individual licenses for each open-source software utilized in this analysis are detailed in the paper analysis code repository.

References

1. Louis, D. N. et al. The 2021 WHO classification of tumors of the central nervous system: a summary. *Neuro Oncol.* **23**, 1231–1251 (2021).
2. Wesseling, P., van den Bent, M. & Perry, A. Oligodendroglioma: pathology, molecular mechanisms and markers. *Acta Neuropathol.* **129**, 809–827 (2015).
3. Zhang, Y. et al. The p53 Pathway in Glioblastoma. *Cancers* **10**, https://doi.org/10.3390/cancers10090297 (2018).
4. Mecca, C., Giambanco, I., Donato, R. & Arcuri, C. Targeting mTOR in glioblastoma: rationale and preclinical/clinical evidence. *Dis. Markers* **2018**, 9230479 (2018).
5. Hu, L. S. et al. Integrated molecular and multiparametric MRI mapping of high-grade glioma identifies regional biologic signatures. *Nat. Commun.* **14**, 6066 (2023).
6. Neftel, C. et al. An integrative model of cellular states, plasticity, and genetics for glioblastoma. *Cell* **178**, 835–849.e821 (2019).
7. Kim, Y. et al. Highly multiplexed spatially resolved proteomic and transcriptional profiling of the glioblastoma microenvironment using archived formalin-fixed paraffin-embedded specimens. *Mod. Pathol.* **36**, 100034 (2023).
8. Ren, Y. et al. Spatial transcriptomics reveals niche-specific enrichment and vulnerabilities of radial glial stem-like cells in malignant gliomas. *Nat. Commun.* **14**, 1028 (2023).
9. Ravi, V. M. et al. Spatially resolved multi-omics deciphers bidirectional tumor-host interdependence in glioblastoma. *Cancer Cell* **40**, 639–655.e613 (2022).
10. De Falco, A., Caruso, F., Su, X. D., Iavarone, A. & Ceccarelli, M. A variational algorithm to detect the clonal copy number substructure of tumors from scRNA-seq data. *Nat. Commun.* **14**, 1074 (2023).
11. Jain, S. et al. Single-cell RNA sequencing and spatial transcriptomics reveal cancer-associated fibroblasts in glioblastoma with protumoral effects. *J. Clin. Investig.* **133**, https://doi.org/10.1172/JCI147087 (2023).
12. Lv, X. et al. Decoding heterogeneous and coordinated tissue architecture in glioblastoma using spatial transcriptomics. *iScience* **27**, 110064 (2024).
13. Greenwald, A. C. et al. Integrative spatial analysis reveals a multilayered organization of glioblastoma. *Cell* **187**, 2485–2501.e2426 (2024).
14. Liu, M. et al. Spatial transcriptomics reveals segregation of tumor cell states in glioblastoma and marked immunosuppression within the perinecrotic niche. *Acta Neuropathol. Commun.* **12**, 64 (2024).
15. Mathur, R. et al. Glioblastoma evolution and heterogeneity from a 3D whole-tumor perspective. *Cell* **187**, 446–463.e416 (2024).
16. Huang, L. E. Impact of CDKN2A/B homozygous deletion on the prognosis and biology of IDH-Mutant Glioma. *Biomedicines* **10**, https://doi.org/10.3390/biomedicines10020246 (2022).
17. Shirahata, M. et al. Novel, improved grading system(s) for IDH-mutant astrocytic gliomas. *Acta Neuropathol.* **136**, 153–166 (2018).
18. Yi, E., Chamorro Gonzalez, R., Henssen, A. G. & Verhaak, R. G. W. Extrachromosomal DNA amplifications in cancer. *Nat. Rev. Genet.* **23**, 760–771 (2022).
19. Von Hoff, D. D. et al. Elimination of extrachromosomally amplified MYC genes from human tumor cells reduces their tumorigenicity. *Proc. Natl. Acad. Sci. USA* **89**, 8165–8169 (1992).
20. Mark, J. & Granberg, I. The chromosomal aberration of double-minutes in three gliomas. *Acta Neuropathol.* **16**, 194–204 (1970).
21. Canute, G. W. et al. The hydroxyurea-induced loss of double-minute chromosomes containing amplified epidermal growth factor receptor genes reduces the tumorigenicity and growth of human glioblastoma multiforme. *Neurosurgery* **42**, 609–616 (1998).
22. Zhou, Y. H. et al. The role of EGFR double minutes in modulating the response of malignant gliomas to radiotherapy. *Oncotarget* **8**, 80853–80868 (2017).

23. Morton, A. R. et al. Functional enhancers shape extrachromosomal oncogene amplifications. *Cell* **179**, 1330–1341.e1313 (2019).
24. Wu, S. et al. Circular ecDNA promotes accessible chromatin and high oncogene expression. *Nature* **575**, 699–703 (2019).
25. Walentynowicz, K. A. et al. Single-cell heterogeneity of EGFR and CDK4 co-amplification is linked to immune infiltration in glioblastoma. *Cell Rep.* **42**, 112235 (2023).
26. Yoshihara, K. et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* **4**, 2612 (2013).
27. Bassiouni, R. et al. Spatial transcriptomic analysis of a diverse patient cohort reveals a conserved architecture in triple-negative breast cancer. *Cancer Res.* **83**, 34–48 (2023).
28. Mazzoleni, A. et al. Chromosomal instability: a key driver in glioma pathogenesis and progression. *Eur. J. Med. Res.* **29**, 451 (2024).
29. Coy, S. et al. Single cell spatial analysis reveals the topology of immunomodulatory purinergic signaling in glioblastoma. *Nat. Commun.* **13**, 4814 (2022).
30. Gao, R. et al. Delineating copy number and clonal substructure in human tumors from single-cell transcriptomes. *Nat. Biotechnol.* **39**, 599–608 (2021).
31. Serin Harmanci, A., Harmanci, A. O. & Zhou, X. CaSpER identifies and visualizes CNV events by integrative analysis of single-cell or bulk RNA-sequencing data. *Nat. Commun.* **11**, 89 (2020).
32. Muller, S., Cho, A., Liu, S. J., Lim, D. A. & Diaz, A. CONICS integrates scRNA-seq with DNA sequencing to map gene expression to tumor sub-clones. *Bioinformatics* **34**, 3217–3219 (2018).
33. Chen, L. et al. STmut: a framework for visualizing somatic alterations in spatial transcriptomics data of cancer. *Genome Biol.* **24**, 273 (2023).
34. Gao, T. et al. Haplotype-aware analysis of somatic copy number variations from single-cell transcriptomes. *Nat. Biotechnol.* https://doi.org/10.1038/s41587-022-01468-y (2022).
35. Erickson, A. et al. Spatially resolved clonal copy number alterations in benign and malignant tissue. *Nature* **608**, 360–367 (2022).
36. An, Z., Aksoy, O., Zheng, T., Fan, Q. W. & Weiss, W. A. Epidermal growth factor receptor and EGFRvIII in glioblastoma: signaling pathways and targeted therapies. *Oncogene* **37**, 1561–1575 (2018).
37. Marallano, V. J. et al. Hypoxia drives shared and distinct transcriptomic changes in two invasive glioma stem cell lines. *Sci. Rep.* **14**, 7246 (2024).
38. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv:1303.3997v2.q-bio.GN* (2013).
39. Van der Auwera, G. O. C., BD. *Genomics in the Cloud: Using Docker, GATK, and WDL in Terra* (O'Reilly Media, 2020).
40. Danecek, P.B. et al. Heng. Twelve years of SAMtools and BCFtools. *GigaScience* **10**, https://doi.org/10.1093/gigascience/giab008 (2021).
41. Cingolani, P. et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly* **6**, 80–92 (2012).
42. Sherry, S. T. et al. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* **29**, 308–311 (2001).
43. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
44. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **14**, 417–419 (2017).
45. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
46. Anders, S., Pyl, P. T. & Huber, W. HTSeq–a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
47. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol* **36**, 411–420 (2018).
48. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol* **20**, 296 (2019).
49. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849 (2016).
50. Harris et al. Array programming with NumPy. *Nature* **585**, 357–362 (2020).
51. McKinney, W. Data structures for statistical computing in python. *SciPy* **445**, 51–56 (2010).
52. Jeffreys, H. *Theory of probability* (Clarendon Press, 1961).
53. Visser, I. S. Maarten. depmixS4: An R Package for Hidden Markov Models. *J. Stat. Softw.* **36**, 1–21 (2010).
54. Peterson, R. A. & Cavanaugh, J. E. Ordered quantile normalization: a semiparametric transformation built for the cross-validation era. *J. Appl. Stat.* **47**, 2312–2327 (2020).
55. Peterson, R. A. Finding optimal normalizing transformations via bestNormalize. *R J.* **13**, 310–329 (2021).
56. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer-Verlag, 2016).
57. Gel, B. & Serra, E. karyoploteR: an R/Bioconductor package to plot customizable genomes displaying arbitrary data. *Bioinformatics* **33**, 3088–3090 (2017).
58. Gel, B. et al. regioneR: an R/Bioconductor package for the association analysis of genomic regions based on permutation tests. *Bioinformatics* **32**, 289–291 (2016).
59. Lawrence, M. et al. Software for computing and annotating genomic ranges. *PLoS Comput. Biol.* **9**, e1003118 (2013).
60. Garnier, S., Ross, Noam, Rudis, Robert, Camargo, Pedro A, Sciaini, Marco, Scherer, Cédric. viridis(Lite) - Colorblind-friendly color maps for R. https://doi.org/10.5281/zenodo.4679423 (2024).
61. Korsunsky, I. et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).

Acknowledgements

We thank the patients for their donations. We acknowledge the USC Norris Comprehensive Cancer Center Translational Pathology Core Facility for cryostat tissue sectioning, Dr. Seth Ruffins and the USC Stem Cell Optical Imaging Facility for microscopy, and the USC Norris Molecular Genomics Core and USC Keck Genomics Platform for sequencing. F.C. was supported by KL2TR001854 from the National Center for Advancing Translational Science (NCATS) of the U.S. National Institutes of Health. The content is solely the authors' responsibility and does not necessarily represent the official view of the National Institutes of Health.

Author contributions

M.G.W. and C.G.M. performed the spatial transcriptomic experimental work. M.G.W. carried out the bioinformatic analysis and tool development. F.C. contributed to the writing of the manuscript and expertise on the topic. B.Z. and J.J.Y.L. contributed to the computational analysis. R.B. assisted with the computational analysis and contributed to the manuscript's writing. N.E.G. III helped with sample selection and collection. K.H. provided pathology expertise and helped guide sample selection. J.D.C. contributed to the project experimental design. G.Z. and D.W.C. supervised and developed the project.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-025-59805-z.

Correspondence and requests for materials should be addressed to Gabriel Zada or David W. Craig.

Peer review information *Nature Communications* thanks Dieter Henrik Heilan, Xiaonan Zhang and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.