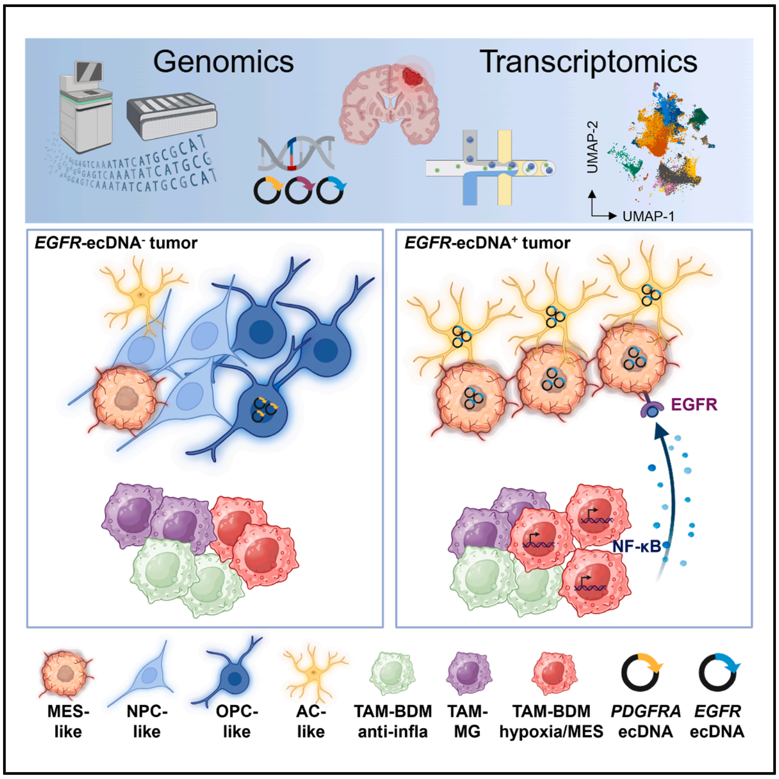
Cell Reports

Article

# Integrated genomic and transcriptomic profiling of glioblastoma reveals ecDNA-driven heterogeneity and microenvironmental reprogramming

## Graphical abstract



## Highlights

- Oncogene-associated ecDNA drives distinct glioblastoma transcriptomic subtypes
- *EGFR* ecDNA structural complexity links to genomic instability and patient survival
- Hypomethylated *EGFR* ecDNA promotes MES- and AC-like tumor phenotypes
- *EGFR* ecDNA shifts TAMs to a metabolically active state and reinforces a signaling loop

## Authors

Wenshu Tang, Wing Lun Lee,
Cario W.S. Lo, ..., Gilberto K.K. Leung,
Aya El Helali, Brian H.Y. Chung

## Correspondence

ahelali@hku.hk (A.E.H.),
bhychung@hku.hk (B.H.Y.C.)

## In brief

Tang et al. provide a comprehensive genomic and transcriptomic characterization of extrachromosomal DNA (ecDNA) in glioblastoma. They reveal that *EGFR* ecDNA shapes transcriptional subtypes and reprograms the tumor microenvironment by stabilizing metabolically active tumor-associated macrophages. These findings uncover mechanistic links between ecDNA architecture and glioblastoma progression, highlighting potential therapeutic vulnerabilities.

Tang et al., 2025, Cell Reports *44*, 116426
October 28, 2025 © 2025 The Authors. Published by Elsevier Inc.
https://doi.org/10.1016/j.celrep.2025.116426

CellPress

Cell Reports

CellPress
OPEN ACCESS

Article

# Integrated genomic and transcriptomic profiling of glioblastoma reveals ecDNA-driven heterogeneity and microenvironmental reprogramming

Wenshu Tang,[1,7] Wing Lun Lee,[3,7] Cario W.S. Lo,[1] Annie T.W. Chu,[1] Karrie Mei-Yee Kiang,[4] Wei Ma,[1] Amy H.Y. Tong,[1] Dingge Ying,[1] Jamie S.L. Kwok,[1] David Shih,[6] Gilberto K.K. Leung,[4,5] Aya El Helali,[3,*] and Brian H.Y. Chung[1,2,8,*]
[1]Hong Kong Genome Institute, Hong Kong SAR, China
[2]Department of Pediatrics and Adolescent Medicine, School of Clinical Medicine, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong SAR, China
[3]Department of Clinical Oncology, School of Clinical Medicine, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong SAR, China
[4]Department of Surgery, School of Clinical Medicine, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong SAR, China
[5]State Key Laboratory of Brain and Cognitive Sciences, The University of Hong Kong, Hong Kong SAR, China
[6]School of Biomedical Science, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong SAR, China
[7]These authors contributed equally
[8]Lead contact
*Correspondence: ahelali@hku.hk (A.E.H.), bhychung@hku.hk (B.H.Y.C.)
https://doi.org/10.1016/j.celrep.2025.116426

## SUMMARY

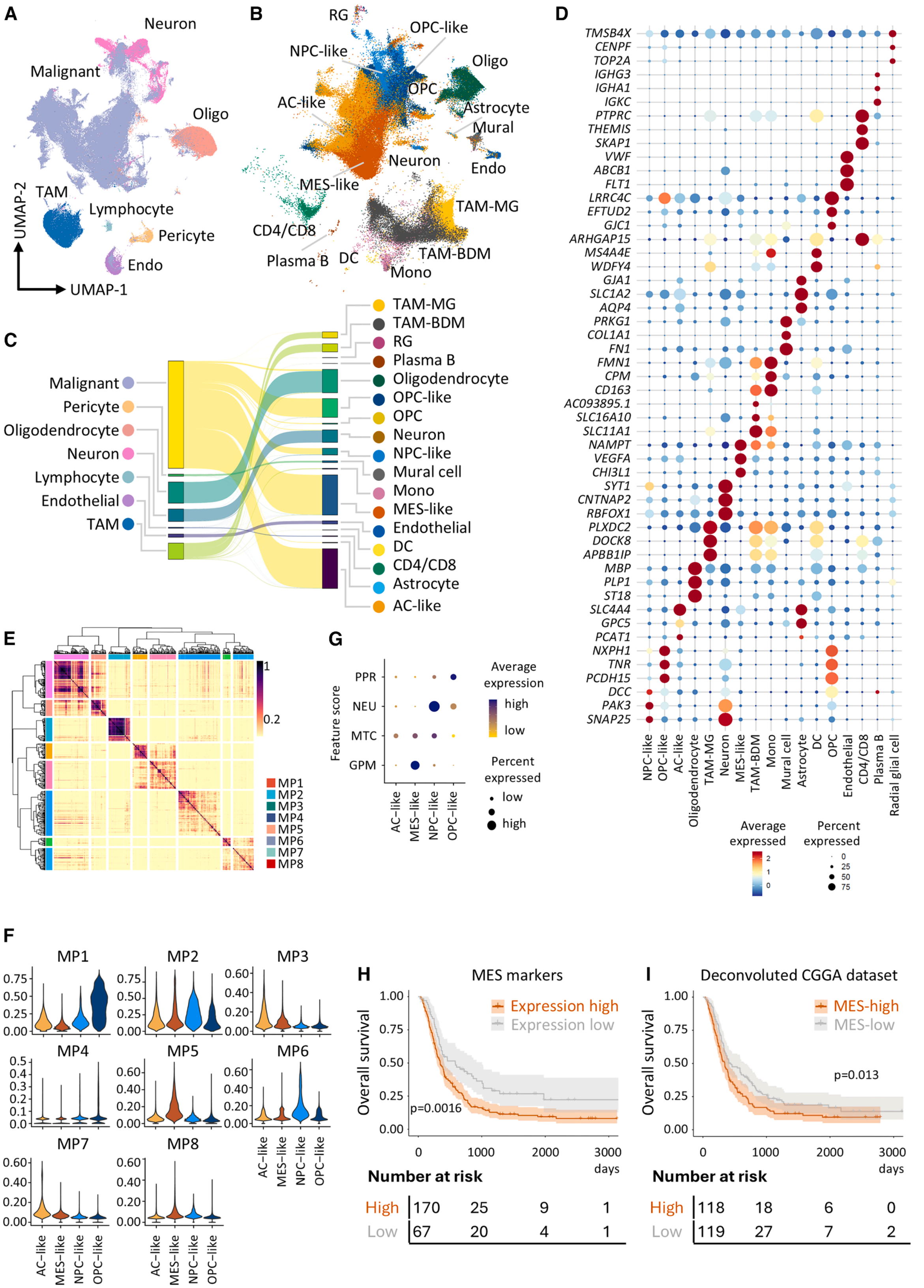
Glioblastoma (GB) is an aggressive brain tumor with limited treatment options, making it crucial to integrate genomic and transcriptomic profiling to identify genetic alterations and cellular functional states. We perform short- and long-read whole-genome sequencing (WGS) and single-nucleus RNA sequencing on 42 Chinese GB patients to characterize the intra- and extrachromosomal mutation landscape, cell-type composition, and pathway activity. Our analysis identifies amplified oncogenes, including *EGFR*, *MYC*, *CDK4*, *PDGFRA*, and *PPARGC1A*, localized on extrachromosomal DNA (ecDNA). Notably, *EGFR* ecDNA harbors distinct structures that correlate with patient survival and exhibit a unique DNA methylation pattern that influences gene expression, driving malignant cell differentiation toward MES-like and AC-like subtypes. Specifically, *EGFR* ecDNA stabilizes tumor-associated macrophages in a hypoxia- and metabolism-driven state, reinforcing a reciprocal AREG-EGFR signaling loop with mesenchymal-like tumor cells. Together, these findings uncover a mechanistic link between ecDNA architecture, transcriptional subtypes, and microenvironmental remodeling, offering critical insights for advancing precision oncology in GB.

## INTRODUCTION

Glioblastoma (GB) is the most common adult malignant brain tumor, and its prognosis remains poor, with an annual diagnosis of 80–100 cases in Hong Kong.[1] Molecular diagnosis involves the evaluation of isocitrate dehydrogenase (IDH) mutation, methylguanine methyltransferase promoter methylation (pMGMT), ATRX chromatin remodeler loss (ATRX LOF), telomerase reverse transcriptase (TERT) promoter mutation, gain in chromosome 7 and loss of chromosome 10 (Chr7+/Chr10−), epidermal growth factor receptor (EGFR) amplification, and cyclin-dependent kinase inhibitor 2A (CDKN2A) deletion.[2] Despite these measures, the prognosis remains poor, with a median overall survival (OS) of approximately 10.6 months.[1] This dismal outcome is largely attributable to the complex genomic landscape and transcriptomic heterogeneity of GB, which contribute to resistance to standard therapies. In this study, we identified a cohort of Chinese patients with histologically and molecularly confirmed IDH-wild-type GB, and we comprehensively characterized the genomes and transcriptomes of these tumors by short- and long-read sequencing.

Recent developments in whole-genome sequencing (WGS) have been crucial for understanding the molecular underpinnings of GB.[3] WGS has provided detailed insights into the genetic features of GB, which have been instrumental in understanding the prognostic biomarkers and mechanisms of treatment resistance, including a consortium of alterations leading to poor prognosis.[4,5]

Extrachromosomal DNA (ecDNA) has garnered increasing attention as a key driver of tumor heterogeneity and therapeutic resistance in various cancer types.[6] Previous studies have linked circular ecDNA to poor prognosis due to its capacity to amplify oncogenes and undergo rapid evolution in response to treatment.[7] Cytogenetic or imaging techniques such as DNA fluorescence *in situ* hybridization (FISH) and optical mapping are employed to visualize ecDNA within cells; however, they rely



*(legend on next page)*

on identifying specific target sequences and present challenges in manipulating metaphase in clinical samples. Conversely, WGS facilitates impartial identification of ecDNA, copy number variations (CNVs), structural variants (SVs), single nucleotide variants (SNVs), and pan-genomic markers such as tumor mutational burden (TMB) and homologous recombination deficiency (HRD).[4] This comprehensive methodology enables the characterization of intricate amplicons, thereby aiding in recognizing potential mutually exclusive and co-occurring driver events and bolstering the rationale for combination therapies. Moreover, long-read WGS using nanopore technology not only enables accurate detection of large and complex structural rearrangements but also provides information on epigenetic features, specifically DNA methylation, which allows a detailed examination of methylation patterns on ecDNA.

Although clinical factors such as age, extent of resection, and genetic factors such as IDH mutations and MGMT promoter methylation status are well-established predictors of patient overall survival, transcriptome profiling provides additional valuable information by reflecting the gene expression patterns and the functional state of different cell populations. *EGFR* amplification is a hallmark of GB, yet its role in shaping the malignant cell subtype and tumor microenvironment (TME), particularly immune and stromal populations, is less well characterized. Tumor-associated macrophages (TAMs) constitute a major component of the GB microenvironment and have been implicated in promoting tumor growth, immune evasion, and resistance to therapy. However, the molecular mechanisms by which ecDNA influences TAM immune-suppressive functions remain unclear. Understanding how *EGFR*-driven tumors reshape the TME and establish reciprocal signaling loops with immune cells is critical for developing targeted therapies that disrupt tumor-supportive crosstalk and enhance anti-tumor immunity.

To address the current knowledge gap, we conducted our study under the precision health theme of the Hong Kong Genome Project (HKGP), a large-scale genome project in Hong Kong supported by the Health Bureau of the Hong Kong SAR Government,[8] to provide a detailed multi-omic view of histologically and molecularly confirmed IDH-wildtype GB. We characterized the single-cell populations within GB tumors and identified complex amplicons, and other genetic alterations in GB. Moreover, we highlighted the distinct molecular tumor subtypes and TAM remodeling processes driven by ecDNA. The identified ecDNAs provide valuable insights into the biology of GB and uncover ecDNA vulnerabilities as potential therapeutic avenues for GB treatment. We thus provide a valuable resource for uncovering the mechanisms of ecDNA-driven tumorigenesis and identifying therapeutic targets against this aggressive disease for which limited effective treatments currently exist.

## RESULTS

### MES-like tumor cells identified by single-nucleus transcriptomics predict poor clinical outcomes

A total of 42 patients with histologically confirmed GB were recruited between March 2021 and July 2022. All 42 patients are IDH1 wild type. Of these, 26 patients were male (61.9%; median age 63 years) and 16 were female (38.1%, median age 55.5 years). The cohort comprised 34 primary and 8 recurrent cases. The median OS was 14.79 months (95% confidence interval [CI]: 8.16–21.41 months). Detailed clinical information is presented in Table S1.

To systematically characterize the cellular heterogeneity and TME of GB, we performed small nuclear RNA sequencing (snRNA-seq) on fresh frozen specimens from 40 patients. Following stringent filtering, we obtained 310,299 cells, with a median of 3,505 genes detected per cell. Dimensionality reduction using uniform manifold approximation and projection (UMAP) revealed seven major cell clusters: malignant cells, oligodendrocytes, neurons, TAMs, endothelial cells, pericytes, and lymphocytes (Figure 1A). To improve the accuracy and robustness of cell type annotation, we mapped our dataset onto GBMap,[9] a well-curated GBM reference atlas comprising over 1.1 million single cells (Figure 1B). The results from this reference-based mapping were highly concordant with our unsupervised clustering (Figure 1C). Among malignant cells, we identified four canonical transcriptional states: astrocyte-like (AC-like), neural precursor cell-like (NPC-like), oligodendrocyte precursor cell-like (OPC-like), and mesenchymal-like (MES-like) states (Figures 1C and 1D). AC-like and MES-like cells were the most prevalent, together accounting for approximately 50% of all malignant cells. AC-like cells were marked by high expression of astrocytic markers (e.g., *SLC4A4*, *PCAT1*), while MES-like cells expressed mesenchymal and angiogenic genes such as *NAMPT*, *VEGFA*, and *CHI3L1* (Figure 1D).

Recognizing the value of pathway-based classification frameworks in understanding GB cellular states,[10] we next sought to dissect the functional programs underlying different transcriptional states. InferCNV analysis confirming that cells annotated

**Figure 1. MES-like tumor cells identified by single-nucleus transcriptomics predict poor clinical outcomes**

(A) UMAP projection of 310,299 single cells isolated from 40 in-house GBM tumor tissues, colored by *de novo* graph-based clustering and inferred cell types.

(B) UMAP projection of the same 310,299 single cells, colored by predicted cell types based on reference-mapping to the GBMap dataset.

(C) Sankey plot illustrating the relationship between graph-based cell clusters and reference-mapping-based cell clusters in the snRNA-seq dataset.

(D) Dot plot showing three canonical markers among the top differentially expressed genes across clusters in the snRNA-seq dataset.

(E) Heatmap of Jaccard similarity depicting robust NMF programs in all malignant cells and their clustering into eight metaprograms.

(F) Violin plots displaying the distribution of metaprogram signature scores across the four malignant cell types.

(G) Dot plot showing the feature scores of four pathway-defined subtypes (PPR, NEU, MTC, GPM) across all malignant cells in the snRNA-seq dataset.

(H) Kaplan-Meier curves depicting overall survival in the CGGA glioma bulk RNA-seq dataset, stratified by high ($n = 170$) and low ($n = 67$) marker scores based on gene TPM. The cut-off point was determined by maximizing the difference between the true positive rate (TPR) and false positive rate (FPR).

(I) Kaplan-Meier survival analysis according to the proportion of MES-like malignant cells in the CGGA bulk RNA-seq dataset, with the median used as the cut-off for stratification. Statistical significance was calculated using a log-rank test in (H and I).

See also Figure S1.

as AC-, MES-, OPC-, or NPC-like malignant cells had significantly higher CNV scores than non-malignant cells (Figures S1A and S1B). We then applied *GeneNMF*,[11] an NMF approach optimized for single-cell data to identify robust, shared transcriptional programs across patients. This method decomposes the gene expression matrix into eight co-activated *metaprograms* (MPs), each representing a distinct set of co-expressed genes associated with specific biological pathways (Figure 1E; Table S2). Clustering and embedding cells based on metaprogram activity revealed clear segregation of malignant states, with each cell type occupying a distinct region in metaprogram-based UMAP space (Figure S1C). This suggests that the major GB cell states are defined not only by transcriptomic differences but also by distinct underlying biological programs. Metaprogram analysis revealed that OPC-like cells primarily activated MP1 and MP4, associated with cell cycle and proliferation pathways (Figures 1F, S1D, and S1E). MES-like cells were enriched for MP5 and MP8, highlighting processes such as epithelial-mesenchymal transition (EMT), hypoxia, and glycolysis, hallmarks of a metabolically reprogrammed, invasive tumor phenotype. NPC-like cells exhibited high activity in MP2 and MP6, reflecting oxidative phosphorylation and synaptic signaling, while AC-like cells activated MP3 and MP7, which were associated with inflammatory signaling, including interleukin-6 (IL-6)-Janus kinase (JAK)-signal transducer and activator of transcription (STAT) pathway (Figures 1F, S1D, and S1E).

To assess concordance with established pathway-based GBM subtypes, we scored each malignant cell using gene signatures defined by Garofano et al.[10]: proliferative/progenitor (PPR), neuronal (NEU), mitochondrial (MTC), and glycolytic/plurimetabolic (GPM) (Figure 1G). MES-like cells had the highest GPM scores, consistent with the hypoxia/glycolysis-enriched metaprograms identified via *GeneNMF*. NPC-like cells exhibited the highest NEU scores, consistent with the synaptic signaling identified in MP6, reaffirming their neurodevelopmental origin. OPC-like cells were most enriched for the PPR signature, reflecting their proliferative characteristics and correspondence with the MP1/MP4 proliferation modules. Together, this dual-layered classification offers a more biologically grounded understanding of GB heterogeneity.

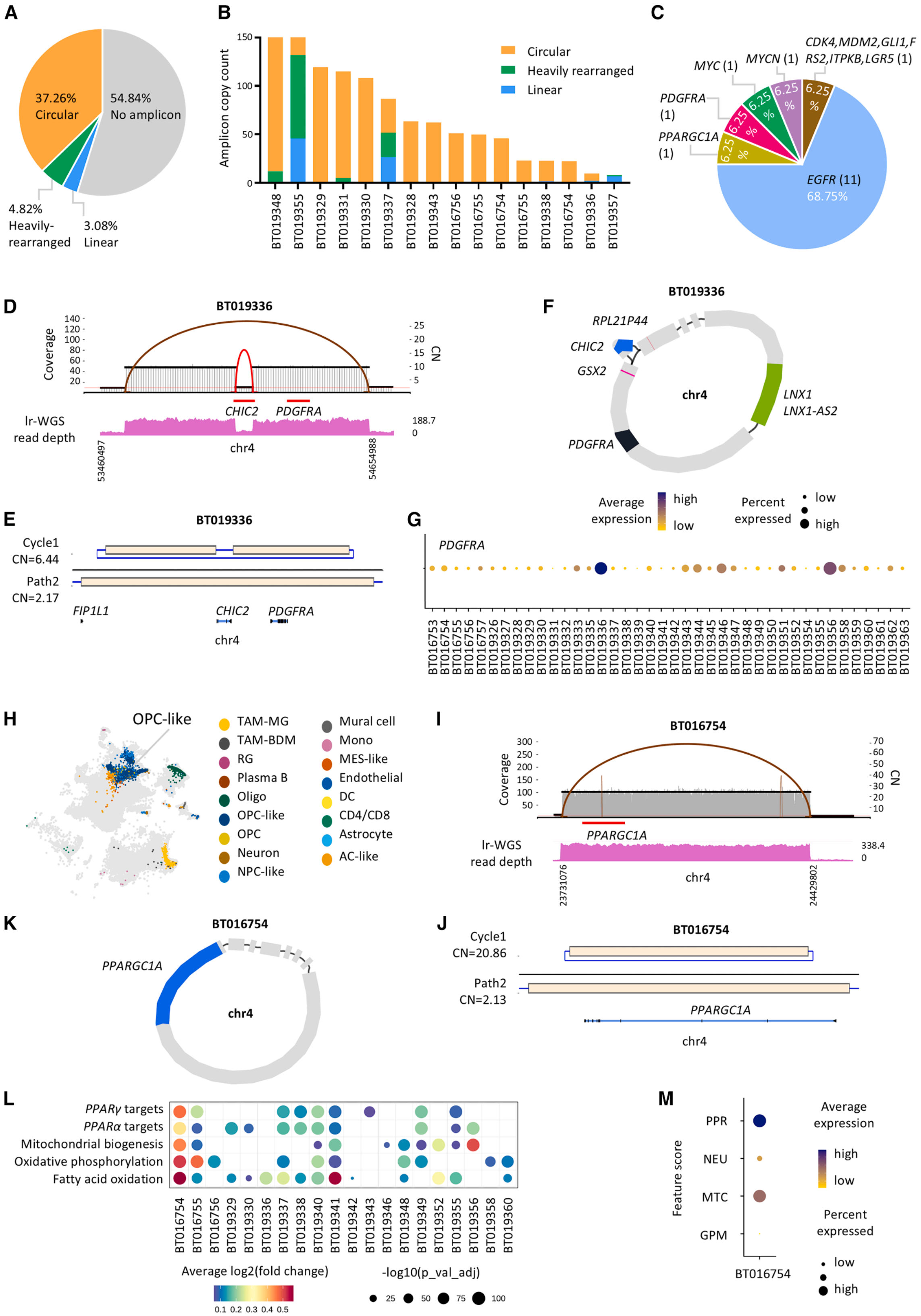
To investigate the clinical relevance of the malignant cell states identified, we assessed the association between their marker genes and OS in GB patients using bulk RNA-seq data from the Chinese Glioma Genome Atlas (CGGA; $n = 237$).[12] Notably, marker genes specific to MES-like malignant cells were significantly associated with poor OS ($p < 0.05$; Figure 1H). Next, we constructed a GBM-specific reference matrix based on the single-nucleus gene expression profiles of each major cell type identified in our dataset (Figure 1B). Using this reference, we applied the CIBERSORTx deconvolution algorithm[13] to infer the cellular composition of bulk CGGA tumors, including both malignant and immune cell populations. This analysis confirmed that a higher proportion of MES-like tumor cells was significantly associated with worse patient survival ($p < 0.05$; Figure 1I). To validate these findings with an orthogonal approach, we applied BayesPrism,[14] a probabilistic deconvolution framework to predict cellular composition from bulk RNA-seq data. BayesPrism independently confirmed the association between MES-like cell states and adverse survival outcomes in the CGGA cohort (Figure S1F). Notably, even within our relatively small patient cohort, we observed that tumors classified as MES- or AC-like subtypes following gross total resection (GTR) exhibited significantly poorer survival than GTR-treated tumors dominated by the proneural subtype ($p = 0.02$; Figure S1G). To further elucidate the molecular subtypes of each patient and assess their risk, we calculated the proportion of each malignant cell subtype per patient (Figure S1H). Notably, six out of eight recurrent samples were dominated by MES-like tumor cell populations, which exhibited the highest enrichment for the GPM signature (Figure S1I). These findings reinforce the notion that MES-like states are not only enriched in recurrent GB but are also associated with poor clinical outcomes, which provides critical insights into patient prognosis and highlights the need for tailored therapeutic strategies for patients with MES subtypes.

### Short- and long-read sequencing uncovers diverse oncogene-carrying ecDNA architectures

We aimed to explore the role of genetic aberrations in fostering transcriptional heterogeneity. To this end, short-read WGS was conducted with tumor and matched germline DNA samples from 42 GB patients, among which 31 pairs of samples with tumor contents greater than 30% were used for downstream analysis. The sequencing results revealed an average sequencing depth of 132.22× for tumor samples and 34.84× for matched germline samples (Table S3). Additionally, long-read WGS analysis was performed on 32 tumors, providing a median read coverage of 98.80× (read length N50 = 10–20 kbp), to obtain the tumor methylation profile and validate the complex rearrangements identified by short-read WGS.

Given the potentially important implication of ecDNA in GB, we examined the presence of ecDNA and its role in regulating tumor differentiation and evolution at the transcriptional level. AmpliconArchitect (AA)[15] was applied to the short-read WGS data to identify the highly amplified regions (copy number [CN] > 4, region size > 10 kbp) and reconstruct the amplicons in each sample. The inferred breakpoints that formed the amplicon structure and the genes involved in ecDNA for each patient were summarized in Table S4. In summary, ecDNA was detected in 14 of 31 (45.16%) GB tumors, predominantly as circular (Figures 2A and 2B), with one sample BT019357 carrying *MYCN* ecDNA shown as linear structure, consistent with previous observations from the PCAWG-glioblastoma study.

To validate the complex architecture of these amplicons, particularly those containing multiple segments, we leveraged ONT long-read sequencing data. ONT's longer reads are better suited to map complex structural variants and resolve rearranged or duplicated regions. We employed two approaches for ecDNA characterization using long-read sequencing data: a reference-mapping-based algorithm named CoRAL,[16] to reconstruct heterogeneous ecDNA structures, which is capable of distinguishing between cyclic and linear architectures; and a *de novo* assembly approach, Flye,[17,18] to resolve complex amplicons. Results from long-read WGS demonstrated strong concordance with short-read WGS-based predictions in all but one sample. For sample BT019342, inferred to harbor a circular *FGFR3-TACC3* amplified amplicon by short-read data, long-read data instead revealed a linear conformation (Figures S2A



*(legend on next page)*

and S2B). The false positive in AA predictions is likely due to a tandem duplication involving *TACC3* and *FGFR3* (Figure S2C), causing low copy number gains, which are frequently observed in GB. These findings highlight the enhanced accuracy of long-read methods in distinguishing between circular and linear DNA structures, which are often challenging to differentiate using short-read data alone.

Using short-read WGS, we identified ten oncogenes amplified on ecDNA across the cohort (Figure 2C). Among these, *PDGFRA* (Platelet-Derived Growth Factor Receptor Alpha) was amplified on ecDNA in patient BT019336, as detected by AmpliconArchitect (Figure 2D). The circular structure of this ecDNA was validated through CoRAL analysis of long-read data (Figure 2E) and reconstructed using *de novo* assembly (Figure 2F). Both short- and long-read data revealed that *CHIC2* has a lower copy number compared to adjacent regions. Long-read data further revealed two forms of amplicon structures: *CHIC2* is absent in amplicon 1 (cycle 1) but present in amplicon 2 (path 2), suggesting structural heterogeneity (Figure 2E). This patient exhibited significantly higher expression levels of *PDGFRA* (Figure 2G). Furthermore, single-nucleus RNA profiling revealed clustering toward the pro-neural subtype for cells harboring *PDGFRA* ecDNA (Figures 2H, S2D, and S2E), underscoring the role of *PDGFRA* ecDNA in driving the differentiation of the pro-neural subtype of malignant cells.

In addition, we identified an enrichment of *PPARGC1A* ecDNA in patient BT016754 using both short-read and long-read data (Figures 2I–2K), with correspondingly elevated *PPARGC1A* expression in this patient (Figure S2F). *PPARGC1A* encodes PPARG Coactivator 1 alpha (PGC1α), a transcriptional coactivator that interacts with transcription factors (TFs) such as PPARγ/PPARα and plays a pivotal role in regulating cellular metabolic pathways such as mitochondrial biogenesis, oxidative phosphorylation, and lipid metabolism.[19] Abnormal PGC1α expression enhances tumor metabolic flexibility, and its suppression has been shown to re-sensitize therapeutic-resistant cancer cells to treatment.[20] In this patient, the tumor was predominantly composed of AC-like malignant cells (Figures S2G and S2H). We further analyzed the signature scores for individual patients and found that patient BT016754 exhibited significantly higher signature scores for PPARγ and PPARα target genes, as well as cellular metabolic pathways compared to other samples (Figure 2L). Moreover, scoring based on PPR, NEU, MTC, and GPM subtypes revealed strong enrichment for both PPR and MTC signatures in this patient (Figure 2M). These findings suggest that ecDNA-mediated *PPARGC1A* amplification may drive increased metabolic activity, providing a potential rationale for therapeutic strategies targeting dysregulated metabolic pathways.

Two other amplicons were identified in this cohort. Patient BT016755 harbored ecDNA containing both *PVT1* (Pvt1 oncogene) and *MYC* (MYC proto-oncogene) (Figures S3A–S3C), resulting in markedly elevated expression of both genes (Figure S3D). The promoter of the long non-coding RNA *PVT1* has been shown to drive the potent expression of *MYC* through enhancer hijacking when both are present on ecDNA, forming a regulatory hub through interaction with the bromodomain and extra-terminal (BET) protein BRD4. This could be a potential genomic target for BET inhibitors, such as JQ1, which preferentially suppresses transcription of ecDNA-derived oncogenes.[21,22] In addition, we identified ecDNA co-amplifying *CDK4* (Cyclin-Dependent Kinase 4) and *MDM2* (MDM2 proto-oncogene) in patient BT019331 (Figures S3E–S3H). CDK4 regulates cell cycle progression, while MDM2 negatively regulates the tumor suppressor p53, highlighting ecDNA as a platform for co-amplifying key oncogenic drivers in GB.

## Complex *EGFR*-ecDNA structures predict poor clinical outcomes

Notably, *EGFR* and its neighboring genes, such as *SEC61G* (SEC61 Translocon Subunit Gamma) were the most frequently amplified genes on ecDNA, identified in 11 out of 31 cases in our cohort (Figure 2C). Most *EGFR* amplicons originated exclusively from chromosome 7 (8 out of 11; Figures 3B and S4A), while some exhibited complex interchromosomal structures, such as an amplicon spanning both chromosomes 5 and 7 in patient BT019337 (3 out of 11; Figure S4D). Long-read sequencing provided critical insights into the proportion and distribution of various *EGFR* ecDNA structures within individual samples. We

**Figure 2. Short- and long-read sequencing uncovers diverse oncogene-carrying ecDNA architectures**

(A) Pie chart showing the distribution of different amplicon types across the patient cohort, categorized as circular, heavily rearranged, linear, or with no amplicon detected.

(B) Amplicon copy counts detected by short-read sequencing for each patient, colored by amplicon categories. Only patients with detectable amplicons are shown.

(C) Top frequently amplified genes detected in ecDNA.

(D and I) Two representative ecDNA amplicons involving *PDGFRA* and *PPARGC1A* genes, detected in two distinct patients (upper). Read depth from long-read sequencing of the corresponding genomic regions (lower). The segmentation is based on coverage and copy number estimates. Horizontal black lines represent segments, with their vertical positions corresponding to estimated copy number. Red arcs indicate orientation-discordant read pairs (e.g., forward-reverse), which suggest deletion-like rearrangements.

(E and J) *PDGFRA* and *PPARGC1A* circular amplicons reconstructed by long-read sequencing. "Cycle" refers to circular DNA structures reconstructed from long-read sequencing data, in which genomic segments form a closed loop. "Path" denotes linear arrangements of genomic segments that may represent either linear amplicons or fragmented ecDNAs. Solid blocks indicate genomic segments, while connecting lines or arcs represent structural rearrangements between them. CN denotes the estimated copy number of the genomic segments.

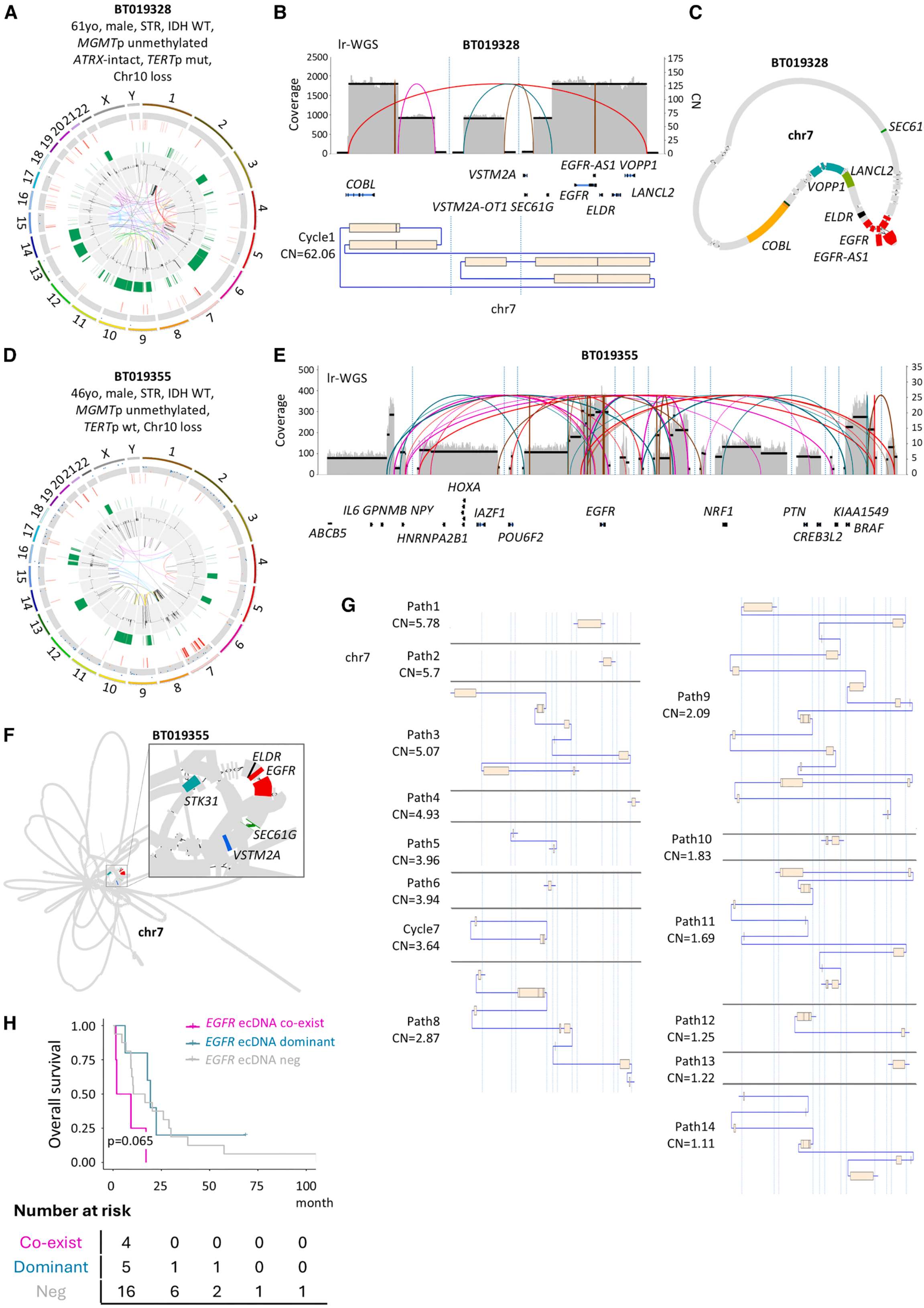
(F and K) *De novo* assembly graphs from long-read sequencing illustrating the positions of amplified genes within the cyclic ecDNA structures.

(G) Dot plots illustrating differential levels of amplified genes *PDGFRA* and *PPARGC1A* in ecDNA. See also Figures S2 and S3.

(H) UMAP of patient BT019336 showing cell clustering toward pro-neuron-like cell types.

(L) Dot plot scoring gene expression for PPARγ and PPARα target genes, as well as three key metabolic pathways in each patient.

(M) Dot plot showing the feature scores of the four pathway-defined subtypes specifically in the patient harboring *PPARGC1A* ecDNA.



(legend on next page)

present molecular profiles of two patients here (Figures 3A and 3D): patient BT019328 was dominated by a single type of circular *EGFR* ecDNA structure, as demonstrated by CoRAL analysis (Figure 3B). This finding was corroborated by *de novo* assembly, which also revealed a clear circular structure (Figure 3C). In contrast, other samples, such as BT019355 (Figure 3E), displayed highly complex rearrangements, as revealed by *de novo* assembly (Figure 3F). Structure disentangling with long-read data uncovered multiple types of *EGFR* ecDNA structures within the same sample (Figure 3G). Additional amplicon profiles illustrating the dominance (Figures S4B and S4C) and co-existence (Figures S4D–S4I) of *EGFR* ecDNA structures are shown in Figure S4 and Table S5.

This observation prompted us to explore whether the presence of multiple *EGFR* ecDNA structure types reflect greater chromosomal dysregulation and its potential association with clinical outcomes. Although our sample size was limited, we found that patients whose tumors harbored multiple co-existing *EGFR* ecDNA structures experienced poor survival outcomes compared to those with a single dominant ecDNA type or very low copy numbers, which showed survival outcomes similar to patients without *EGFR* ecDNA ($p = 0.065$, Figure 3H). These findings emphasize the prognostic importance of ecDNA complexity and the need to differentiate *EGFR* ecDNA structures when assessing patient prognosis.

## *EGFR* ecDNA co-occurs with *EGFRvIII* and is associated with low homologous recombination deficiency scores

We next explored whether intrachromosomal genetic alterations are associated with extrachromosomal genomic events. Interestingly, *EGFR* mutations, including SNVs and the *EGFR*vIII variant, were more frequently observed in *EGFR* ecDNA-positive tumors, with *EGFR*vIII almost exclusively found in *EGFR* ecDNA-positive tumors. This supports the notion that *EGFR*vIII-bearing ecDNA structures may confer a selective advantage under therapeutic pressure (Figure 4A).

Chromosomal arm-level alterations, such as the gain of chromosome 7 ($n = 17$) and loss of chromosome 10 ($n = 18$), were frequently observed in our cohort (Figure S5A). Overexpression and downregulation of multiple genes on chromosomes 7 and 10 were confirmed through InferCNV analysis of the snRNA-seq data (Figure S5B). Using the GISTIC2[23] algorithm to identify recurrent sites of DNA copy number alterations in 31 samples, 46 amplification peaks and 37 deletion peaks were identified (false discovery rate [FDR] < 0.25) (Figure S5D; Table S6). The most common amplification events were at 7p11.2 (*EGFR*), 12q14.1 (*CDK4*), and 4q12 (*PDGFRA*), which was consistent with the presence of ecDNA. Patients with ecDNA exhibit markedly elevated copy numbers of gene segments enriched in ecDNA, with 7p11.2 (*EGFR*) ranging from 22 to 166, 12q13.3–q14.1 (*CDK4-MDM2*) at 30, 4p15.2 (*PPARGC1A*) at 27, and 4q12 (*PDGFRA*) at 13.

HRD and aneuploidy are both known to propagate genomic instability by failing to repair DNA breaks and disrupting normal cell cycle regulation, leading to mutation accumulation and fostering tumor evolution.[24,25] In this cohort, 4 out of 31 samples exhibited genome-wide aneuploidy, with estimated ploidy ranging from 3.39 to 6.71 (Figure S5C; Table S7), and frequent focal deletions at 9p21.3 (*CDKN2A/B*) and 10q23.31 (*PTEN*) (Figure S5D).

Since DNA damage repair (DDR) pathway plays an important role in preserving human genomic stability, we further explored how the copy number alterations affect genes involved in the DDR pathway, such as *ATM*, *BRCA1*, *BRCA2*, and *CHEK2*. We analyzed the number of affected DDR genes in each patient using a comprehensive gene list containing 280 DDR genes from a previous study (Figure S5E).[26] Expression changes were then examined based on each patient's specific set of deleted DDR genes (Table S8). We found that gene expression levels were strongly associated with CNV deletions, with each patient consistently showing markedly reduced expression of the DDR genes affected by copy number loss. In most cases, the deleted genes ranked among the lowest expressed DDR genes within that patient, supporting a direct link between genomic deletion and transcriptional downregulation (Figure S5F). Patients with deletions affecting a higher number of DDR genes showed a higher HRD score, as determined by the loss of heterozygosity, telomeric-allelic imbalance, and large-scale state transition scores (Figure S5G; Table S9). Interestingly, we observed an inverse association between ecDNA presence and HRD score, suggesting that ecDNA maintenance may be favored in genomically stable backgrounds. Furthermore, patients with tumors dominated by a single type of *EGFR* ecDNA exhibited the lowest HRD scores compared to those with multiple co-existing *EGFR* ecDNA structures ($p < 0.05$; Figure 4B).

Genome-wide mutational signature analysis revealed nine distinct single-base substitutions (SBSs) and four small insertion-and-deletion (ID) signatures (Figures 4C and 4D). The most prevalent signature was SBS5, which was typically linked to the aging process (Figure 4C). In comparison to COSMICv3.4,

**Figure 3. Complex *EGFR*-ecDNA structures predict poor clinical outcomes**

(A and D) Representative Circos plots displaying the genome-wide mutational landscape of BT019328 (A) and BT019355 (D). From the outermost to innermost circles: somatic exonic SNVs, copy number gains (red), loss of heterozygosity (gray), copy number losses (green), copy number changes represented as histograms, and structural variants in the central line.

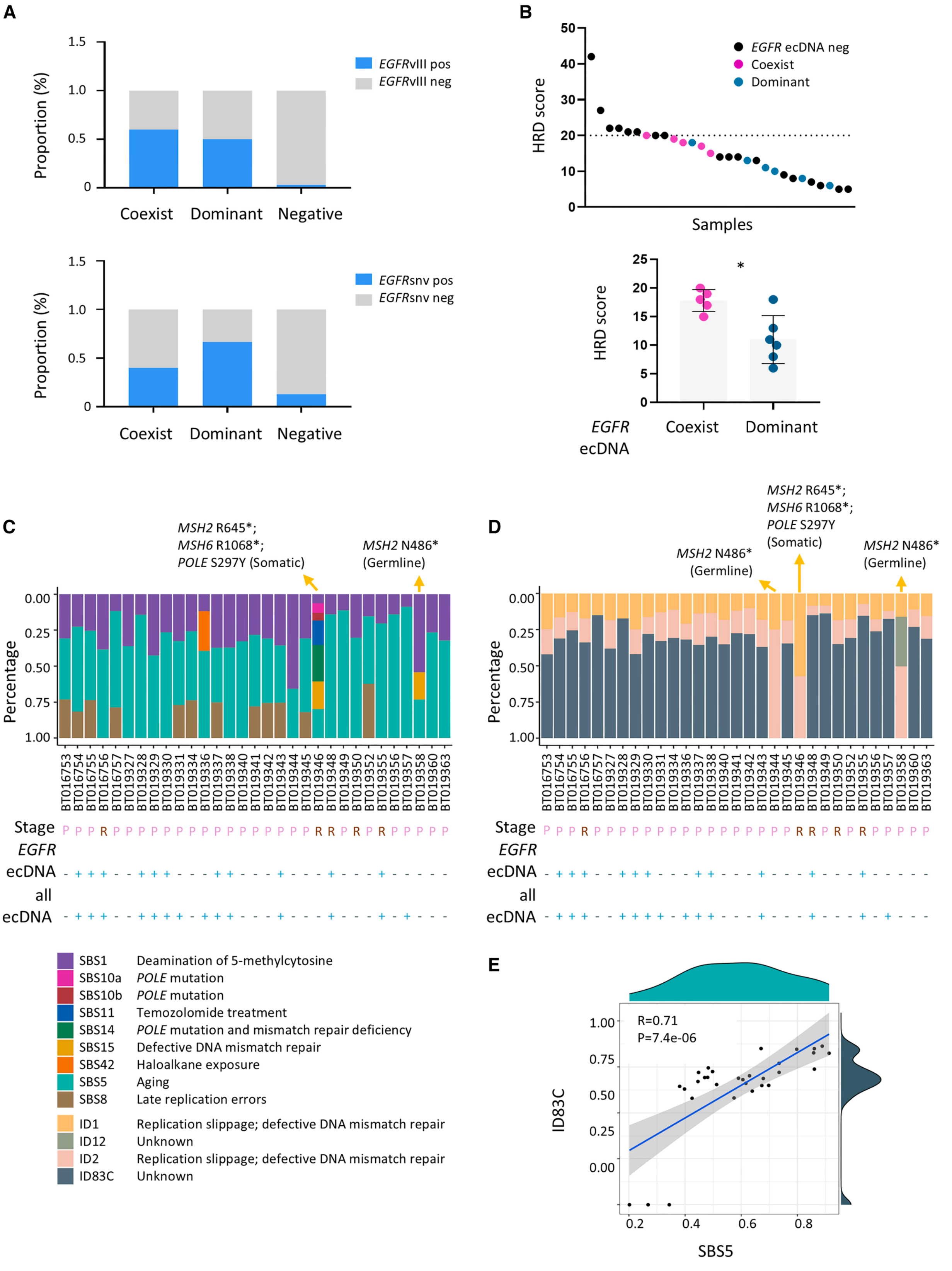
(B and E) The SV view of breakpoints and reconstructed amplicon structures of ecDNA amplicons involving *EGFR* in BT019328 and BT019355 by CoRAL. Arcs represent discordant read pair clusters, colored by orientation: red, deletion-like (length-discordant in expected orientation); brown: duplication-like (everted read pairs); teal, left inversion-like; magenta, right inversion-like. The thickness of each arc qualitatively reflects the amount of supporting paired-end reads.

(C and F) *De novo* assembly graphs from long-read sequencing illustrating the positions of amplified genes within the ecDNA structures.

(G) Reconstructed cyclic ecDNA structure in BT019355 by CoRAL.

(H) Kaplan-Meier survival curves stratified by patients with or without multiple ecDNA structures versus those with a single dominant ecDNA structure. The numbers at risk table indicates how many patients in each classification group (co-exist, dominant, and neg) remained at risk of the death event at specific time points (0, 20, 50, 75, and 100 months). Statistical significance was calculated using a log-rank test.

See also Figure S4.



(legend on next page)

an unreported ID signature (ID83C) was observed, featuring 1 bp deletions and >5 bp insertions at repeats (Figure 4D). ID83C was strongly associated with SBS5, suggesting its potential relevance in aging ($p < 0.05$; Figure 4E). SBS8, which has been suggested to be related to late replication errors based on recent evidence,[27] was more frequent in primary cases than in recurrent cases (Figure 4C). SBS11 associated with temozolomide (TMZ) treatment was observed in one hypermutated recurrent tumor sample from patient BT019346, who received 18 cycles of adjuvant TMZ. Mutational signatures related to *POLE* mutations (SBS10a, SBS10b, and SBS14) were detected in patient BT019346, who harbored an oncogenic *POLE* S297Y somatic mutation. Additionally, signatures associated with defective DNA mismatch repair, including SBS15, ID1, and ID2, were found in three cases (BT019344, BT019346, and BT019358), all of which displayed *MSH2* mutations (Figure 4D). Notably, two patients (BT019344, BT019358) carried the germline pathogenic variant *MSH2* N486*. Both have a family history of colon cancer, suggesting a potential link to cancer predisposition Lynch syndrome and indicating potential responsiveness to immunotherapy.[28] Interestingly, tumors with *MSH2*/*MSH6*/*POLE* mutations exhibiting mismatch repair deficiency signatures seldom contain ecDNA, implying distinct mechanisms of tumor evolution.

### *EGFR* ecDNA is associated with hypomethylation region and facilitates transformation to MES- and AC-like tumor subtypes

To discern the epigenetic landscape in patients with and without *EGFR* ecDNA, we analyzed DNA methylation profiles by assessing 5mCpG signal from long-read WGS data. Interestingly, an open chromatin region in intron 1 of *EGFR*, evidenced by hypomethylated CpG sites, was consistently observed in all patients harboring *EGFR* ecDNA, but not in their counterparts (Figures 5A–5C). While a significantly lower frequency of 5mCpG was observed in patients with *EGFR* ecDNA compared to those without ($p < 0.05$), no significant differences were detected between patients with coexistent or dominant *EGFR* ecDNA (Figure 5B). Leveraging a public single-cell assay for transposase accessible chromatin using sequencing (scATAC-seq) dataset from GB patients,[29] we observed pronounced differences in ATAC-seq signal across the *EGFR* locus when stratifying patients by *EGFR* copy number, using a threshold of six copies (Figure 5C). Additionally, this hypomethylated region not only overlaps with ATAC-seq peaks but also with H3K27ac enrichment signal from another GB cohort,[30] suggesting its role as an accessible *cis*-regulatory element that may drive transcriptional activation of *EGFR* (Figure 5C). Consistently, snRNA-seq data revealed significantly higher expression levels of *EGFR* and its neighboring gene *SEC61G* in patients with *EGFR* ecDNA (Figure 5D). Notably, *EGFR* expression was the highest in AC-like and MES-like cells (Figure S6A). Furthermore, analysis of the TME revealed that patients with *EGFR* ecDNA exhibited transcriptional profiles enriched for MES-like or AC-like signatures or a combination of both compared to those without *EGFR* ecDNA (Figures 5E and S6B). These findings suggest a potential regulatory effect of *EGFR* ecDNA on determining the transcriptome of the tumor cells.

To understand the transcriptional regulatory effect of *EGFR* ecDNA, integrative analysis of *EGFR* ecDNA, methylation, and gene expression were performed. First, we computed a high-dimensional gene expression network using weighted correlation network analysis (hdWGCNA) to identify modules of co-expressed genes in this GB dataset (Figure S6C). Next, these modules were correlated with four traits: the presence of *EGFR* ecDNA, the copy number of *EGFR*, methylation score of the previously identified region in *EGFR* intron 1, and *EGFR* expression level. By quantifying the correlation among these traits and the gene expression signatures derived from individual modules, we identified three differentially expressed modules (Figures 5F and S6D) that were positively correlated with *EGFR* ecDNA, *EGFR* copy number, and *EGFR* expression while being negatively correlated with the methylation level in *EGFR* intron 1: module purple, magenta, and tan (Figure 5G).

Notably, Gene Ontology analysis revealed that the purple module was associated with EMT, tumor necrosis factor alpha (TNF-$\alpha$) signaling via nuclear factor $\kappa$B (NF-$\kappa$B), and angiogenesis, and the magenta module was linked to an inflammatory response (Figure 5H). Furthermore, the top hub genes within the purple and magenta modules, which are highly connected genes within each module, were found to be significantly correlated with worse survival in patients from CGGA ($p < 0.05$; Figure 5I). We next projected module signatures onto the snRNA-seq dataset and observed significant enrichment of the purple and magenta expression modules in MES-like and AC-like cell lineages (Figure 5J). MES-like malignant cells from *EGFR* ecDNA-positive patients exhibited the strongest activation of EMT and TNF-$\alpha$ signaling pathways compared to *EGFR* ecDNA-negative patients (Figure 5K). Interestingly, although the abundance of TAM-BDM (bone marrow-derived macrophages) and TAM-MG (microglia) did not differ significantly between *EGFR* ecDNA-positive and -negative patients (Figure S6B), marked transcriptomic differences were observed. NF-$\kappa$B signaling, a key downstream effector of TNF-$\alpha$, regulates the expression of various targets, including pro-inflammatory cytokines, chemokines, and cell adhesion molecules. To evaluate NF-$\kappa$B pathway activity, we assessed the expression scores based on 46 cytokine or chemokine markers and 11 cell

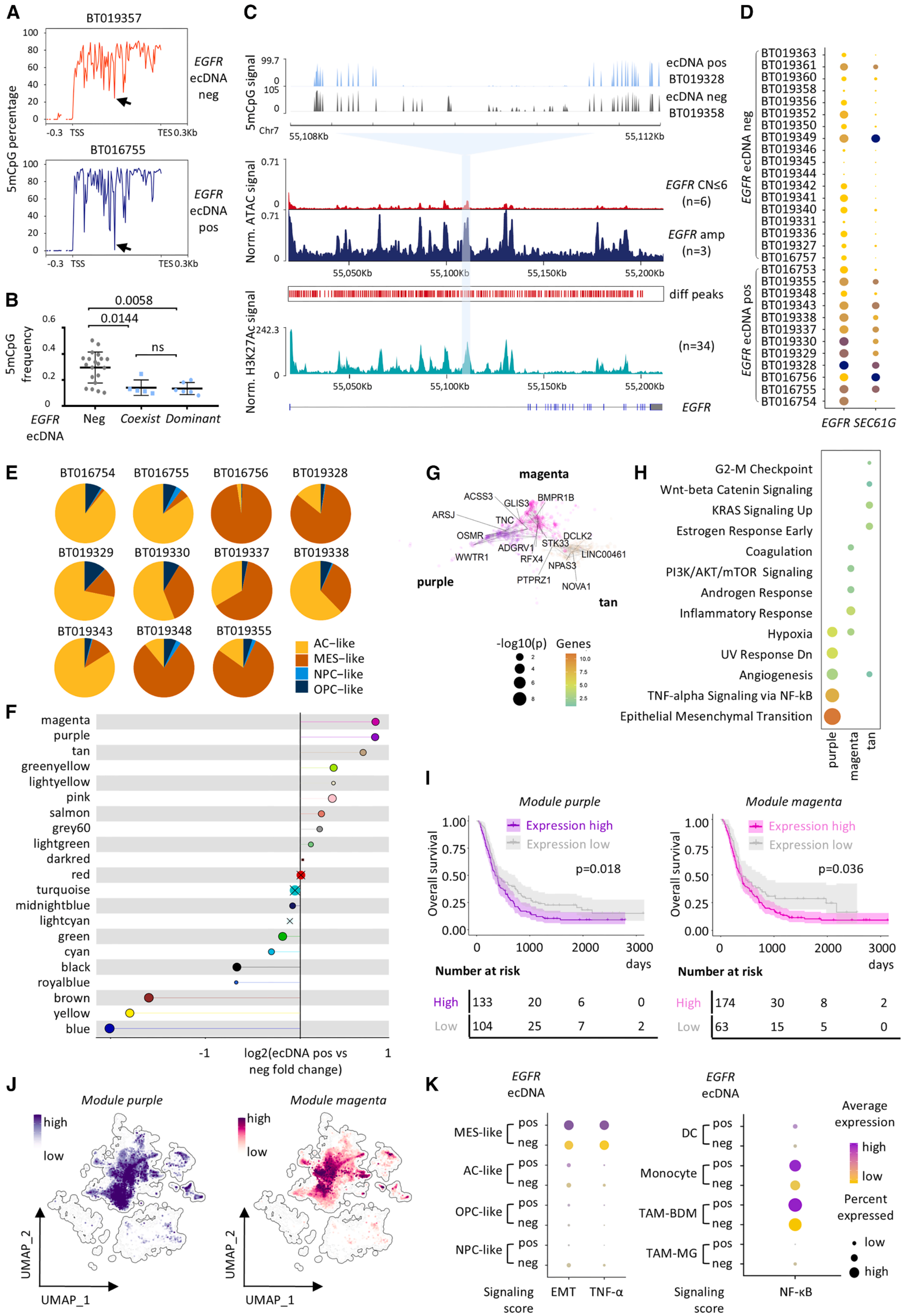
**Figure 4. *EGFR* ecDNA co-occurs with *EGFRvIII* and is associated with low homologous recombination deficiency scores**

(A) Bar plot showing the proportion of *EGFRvIII* and other *EGFR* mutations among patients with or without *EGFR* ecDNA.

(B) Top: dot plot showing HRD score for each patient, colored by *EGFR* ecDNA status. Bottom: HRD score difference between multiple ecDNA structures versus those with a single dominant ecDNA structure. Data are represented as mean ± SD.

(C and D) Upper panels: Composition of SBS (C) and ID (D) mutational signatures in each patient. Lower panels: Stage (P-primary, R-recurrent), *EGFR* ecDNA status and all ecDNA status of each patient.

(E) Correlation between SBS5 and ID83C mutational signatures. Statistical significance was calculated using the Wilcoxon rank-sum test in (B) and Spearman's correlation coefficient test in (E). *$p < 0.05$. See also Figure S5.



(legend on next page)

adhesion molecule markers that are regulated by NF-κB in different subtypes of myeloid lineage cells. Notably, NF-κB signaling exhibited the highest activation in TAM-BDM from *EGFR* ecDNA-positive patients compared to their *EGFR* ecDNA-negative counterparts (Figure 5K). These findings underscore the potential role of ecDNA-bearing tumors in shaping the TME.

### *EGFR* ecDNA reprograms TAM-BDM toward hypoxia/MES differentiation and establishes reciprocal signaling in positive tumors

To further elucidate the role of *EGFR* ecDNA in modulating the TME, we focused on myeloid cell populations and categorized them into five major subtypes: dendritic cells (DCs), monocytes, TAM-BDM (anti-inflammatory), TAM-BDM (hypoxia/MES), and TAM-microglia (MG) (Figure 6A). We then examined the cellular dynamics of these myeloid populations between *EGFR* ecDNA-positive versus -negative tumors. RNA velocity analysis was performed to capture dynamic states of myeloid cells, followed by random walk simulation modeling starting from monocytes, which are known to serve as precursors for bone-marrow-derived macrophages. Notably, in *EGFR* ecDNA-negative tumors, terminal cell states were distributed across multiple myeloid subtypes, whereas in *EGFR* ecDNA-positive patients, terminal cells were predominantly enriched in the TAM-BDM (hypoxia/MES) population, suggesting a lineage bias toward stably differentiated cells (Figure 6B).

To further resolve these dynamic transitions, we applied CellRank2[31] to infer macrostates, defined as clusters of cells exhibiting similar differentiation trajectories. Macrostates were identified using Schur decomposition of the transition matrix and classified as initial, intermediate, or terminal states based on transition probabilities, with the state with highest incoming transition probabilities automatically assigned as the terminal state. Visualization of fate probabilities per terminal state revealed that monocytes predominantly occupied the naive state, positioned centrally within the transition circle (Figure 6C). Cells from *EGFR* ecDNA-negative tumors exhibited a heterogeneous transition profile with four distinct terminal macrostates: TAM-MG, TAM-BDM (anti-inflammatory 1 and 3), and TAM-BDM (hypoxia/MES) (Figure 6C, left). In contrast, *EGFR* ecDNA-positive tumors displayed three terminal macrostates: TAM-BDM (hypoxia/MES), TAM-BDM (anti-inflammatory), and TAM-MG (Figure 6C, right). Importantly, in the *EGFR* ecDNA-positive group, myeloid cells exhibited a higher long-term probability of transitioning and residing in the TAM-BDM (hypoxia/MES) state, as shown by a stationary distribution score of 0.86 (Figure 6E). This was in stark contrast to the *EGFR* ecDNA-negative group, which retained a more diverse transition potential across multiple myeloid subtypes (Figure 6D).

Given the preferential differentiation and stabilization of TAM-BDM (hypoxia/MES) cells in *EGFR* ecDNA-positive tumors, we further investigated their transcriptional signatures. Notably, this cell population displayed upregulation of metabolic pathways, including hypoxia signaling, mTORC1 activation, and glycolysis, alongside TNF-α-NF-κB signaling as shown earlier (Figures 5K and 6F). Additionally, this subset exhibited the highest scores when assessed with signatures from recently reported detrimental myeloid cell populations, including early myeloid-derived suppressor cells (E-MDSCs)[32] and tumor-associated foam cells (TAFs)[33] in GB, characterized by enhanced metabolic and hypoxic activity, tumor-promoting properties, and immunosuppressive effects (Figure 6G).

Latent time analysis further highlighted distinct transcriptional dynamics in TAM-BDM (hypoxia/MES) cells from *EGFR* ecDNA-positive tumors, marked by increased expression of NF-κB-regulated cytokines and chemokines, including *ANGPTL4*, *CXCL8*, *CCL20*, *CXCL2*, *IL1RN*, *CD44*, *FN1*, *IL-6*, and *IL-10* alongside differentiation trajectory (Figure 6H). These findings suggest that *EGFR* ecDNA might reprogram TAM-BDM cells toward an immunosuppressive state via NF-κB-driven transcriptional activation.

To assess potential signaling interactions between TAM-BDM (hypoxia/MES) cells and tumor cells, we applied CellPhoneDB[34] and identified a selective enrichment of *AREG-EGFR* signaling between TAM-BDM (hypoxia/MES) cells and MES-like tumor cells in *EGFR* ecDNA-positive tumors ($p < 0.05$). This signaling axis, which has been reported to

**Figure 5. *EGFR* ecDNA characterization by integrated epigenome and transcriptome profiling**

(A) Methylation profiles across the *EGFR* locus, ranging from the transcription start sites (TSS) to the transcription end sites (TES), in four GB patients. The arrow highlights a peak indicating a differentially methylated region between patients with and without *EGFR* ecDNA.

(B) Dot plot illustrating the differential levels of 5-methylcytosine (5mCpG) between patients harboring *EGFR* ecDNA and those without.

(C) Top: representative 5mCpG methylation profile of the region indicated in (E) from two patients. Middle: genome track of aggregated scATAC-seq data around the EGFR locus, normalized by ReadsInTSS. The tracks are colored by *EGFR* copy number (CN $\leq$ 6, 6 patients, red; CN $\geq$ 38, 3 patients, dark blue). Differentially enriched peaks (FDR $\leq$ 0.000001 and log2FC $\geq$ 3) in patients with *EGFR* amplification are highlighted by red lines below the track. Lower: genome track of aggregated H3K27ac histone modification signals normalized to reads per million from 34 GB patients, spanning all exons and introns of *EGFR*.

(D) Dot plot showing the expression levels of *EGFR* and *SEC61G* across patients in the GB snRNA-seq dataset.

(E) Pie chart illustrating the cellular composition of AC-like, MES-like, NPC-like, and OPC-like cells in all malignant cells from each patient who was positive for *EGFR* ecDNA.

(F) Differential co-expression analysis of module eigengenes reveals modules that are upregulated or downregulated in patients with or without *EGFR* ecDNA.

(G) UMAP visualization of three key gene expression modules (magenta, purple, and tan) that showed significant associations with the presence of *EGFR* ecDNA.

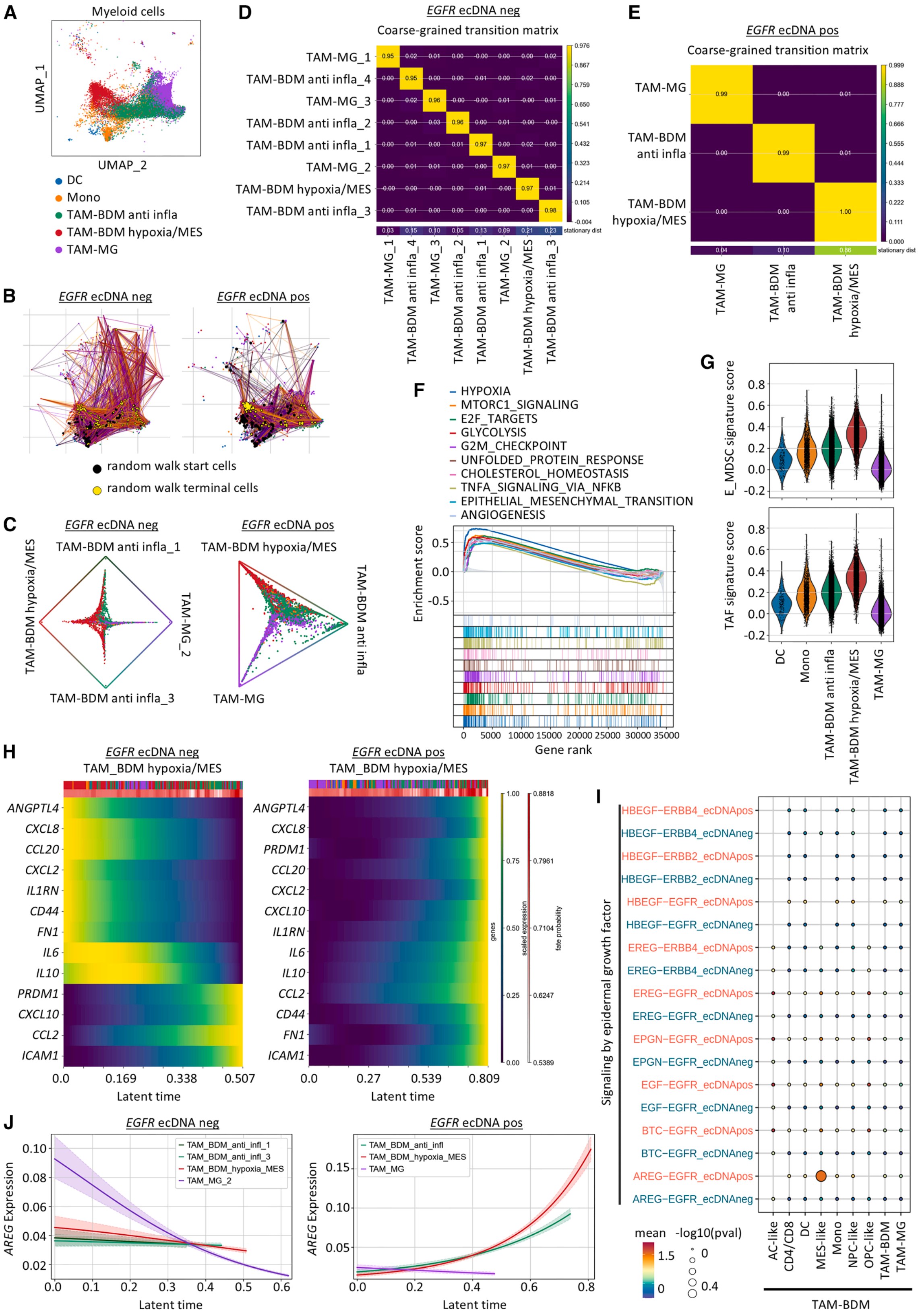
(H) Gene Ontology analysis of the top 100 genes enriched within the three identified gene expression modules.

(I) Kaplan-Meier survival curves stratified by the expression of the top 10 genes in each module, calculated based on gene TPM values in CGGA glioblastoma patients. The cut-off is selected using the point that maximizes the difference between TPR and FPR.

(J) UMAP visualization of purple and magenta eigengene expression in the snRNA-seq dataset.

(K) Dot plot elucidating the expression activity of EMT (200 genes), TNF-α (200 genes), and NF-κB signaling (57 genes). Statistical significance was calculated using a log-rank test in (I).

See also Figure S6.



*(legend on next page)*

promote glioma invasion,[35] was absent in *EGFR* ecDNA-negative tumors, suggesting a potential feedforward loop wherein TAM-BDM (hypoxia/MES) cells reinforce mesenchymal tumor states via *EGFR* activation (Figure 6I). Consistently, *AREG* expression within TAM-BDM (hypoxia/MES) cells exhibited dynamic regulation along the differentiation trajectory, supporting their role as a source of *EGFR* ligands (Figure 6J). Together, these findings indicate that *EGFR* ecDNA not only skews myeloid cell fate toward a hypoxia/metabolism-driven phenotype but also establishes a reciprocal signaling axis with MES-like tumor cells via *AREG-EGFR* crosstalk, ultimately fostering an aggressive and immunosuppressive TME.

## DISCUSSION

This study presents a comprehensive exploration of the genomic and transcriptomic landscape of GB. We elucidate the complex structures of ecDNA through both short- and long-read sequencing, uncovering a unique DNA methylation profile within *EGFR* ecDNA. This methylation pattern and ecDNA are linked to gene expression programs that drive cell differentiation into specific transcriptomic subtypes, including MES-like and AC-like tumor cells. Furthermore, our analysis uncovers a previously unrecognized role of *EGFR* ecDNA in reprogramming the TME by modulating TAM-BDM (hypoxia/MES) function through elevated NF-κB-regulated cytokine and chemokine expression, highlighting a potential mechanism of ecDNA-driven tumor progression and immune evasion.

Leveraging the snRNA-seq dataset, we performed a detailed characterization of GB cell states through gene- and pathway-based clustering. Our results consistently identified MES-like malignant cells as the subtype most strongly associated with poor overall survival, as supported by both single-cell profiles and deconvolution analysis of bulk RNA-seq data from the CGGA cohort. By applying *GeneNMF* across multiple patients, we identified metaprograms that reflect recurrent biological pathways, highlighting core functional modules that define distinct cellular states. MES-like cells were enriched for hallmark processes including EMT, hypoxia, and glycolysis, which are features of a metabolically reprogrammed, invasive tumor phenotype. These cells also exhibited the highest GPM scores as defined by Garofano et al.[10] Notably, we observed a shift toward reduced PPR and increased GPM states in recurrent tumors, suggesting dynamic transcriptional reprogramming associated with recurrence and therapeutic resistance.

We identified an association between the *EGFR* ecDNA and specific transcriptional subtypes. Gene expression modules that significantly correlate with the presence of *EGFR* ecDNA, hypomethylation in intron 1 of *EGFR*, and increased *EGFR* copy number are highly expressed in MES- or AC-like malignant cells. Moreover, our findings revealed a previously unrecognized role of *EGFR* ecDNA in shaping the GB microenvironment by driving a preferential transition of metabolically active TAM-BDM (hypoxia/MES) and establishing a reciprocal signaling loop with mesenchymal-like tumor cells. The observed transcriptomic differences in TAMs between *EGFR* ecDNA-positive and -negative patients provide insights into the immunoregulatory role of ecDNA. TAM-BDM cells are known for their immunosuppressive role in promoting tumor progression.[36,37] Through fate-mapping analyses, we observed that *EGFR* ecDNA-positive tumors exhibit a distinct myeloid cell landscape, with TAM-BDM (hypoxia/MES) cells emerging as the predominant terminal state. These results suggest that *EGFR* ecDNA play a dual role in tumor and immune modulation, not only driving tumor-intrinsic transcriptional programs but also promoting an immunosuppressive TME.

Interestingly, we observed that the presence of ecDNA is nearly mutually exclusive of certain genetic alterations. ecDNA is almost absent in tumors with a high number of DDR gene deletions, hypermutated tumors, and mismatch repair deficiency (MMRd) tumors carrying mutations in *MSH2*/*MSH6* and *POLE*. This observation aligns with findings from a recent study, which revealed an inverse correlation between ecDNA and hypermutation, with MMRd or DNA polymerase deficiency (POLD1/POLEd) signatures being more negatively associated with ecDNA in hypermutated tumors.[38] On the other hand, we observed a reverse association between the presence of ecDNA and a high number of DDR-related gene deletions. The biogenesis of ecDNA is thought to be a follow-up event from double-strand breaks (DSBs) in tumor cells but comparably rare in normal cells, potentially arising from mechanisms like chromothripsis and breakage-fusion-bridge (BFB) cycles.[6] Given the role of DSBs in ecDNA formation, the DDR pathway may provide evidence of ecDNA maintenance in tumor cells.[39] This suggests that DDR/MMR genes may play a role in conferring positive selection preference over replication and progression stress during the tumor cell cycle, potentially influencing ecDNA prevalence. For instance, a functional DDR/MMR gene may allow ecDNA-containing tumors to evade from cell cycle checkpoints and promotes higher heterogeneity coupled with uneven segregation.

**Figure 6. *EGFR* ecDNA reprograms TAM-BDM toward hypoxia/MES differentiation and establishes reciprocal signaling in positive tumors**

(A) UMAP projection of myeloid cell subtypes, colored by graph-based clustering.

(B) UMAP visualization of the differentiation hierarchy of myeloid cells in *EGFR* ecDNA-negative and -positive patients. Black and yellow dots indicate the starting and terminal cells of random walks, respectively.

(C) Circular projections of myeloid cells, depicting fate probabilities toward terminal macrostates. Cells biased toward a specific fate are positioned near the corresponding corner, while naive cells are centered.

(D and E) Heatmap illustrating the classification of macrostates as either initial or terminal, with transitions visualized based on the macrostate-level transition matrix. Macrostates with high diagonal values and strong stationary distribution are defined as stable terminal states.

(F) Gene set enrichment analysis (GSEA) highlighting pathways enriched in marker genes of the TAM-BDM (hypoxia/MES) population.

(G) Violin plot displaying the signature scores of E-MDSC ($n$ = 33 genes) and TAF cells ($n$ = 54 genes) across five myeloid cell subsets.

(H) Heatmap showing gene expression trends of NF-κB-regulated cytokines and chemokines, ordered by latent time.

(I) Dot plot depicting significantly enriched ligand-receptor interactions in the *EGFR* signaling pathway between *EGFR* ecDNA-negative and -positive groups.

(J) Trajectory-specific dynamics of *AREG* in *EGFR* ecDNA-negative and -positive groups, colored by terminal macrostates.

Our findings revealed that ecDNA is almost absent in patients with relatively higher HRD scores; this aligns with recent retrospective machine-learning studies in pan-cancer analyses, which demonstrated the upregulation of DSB-related DDR genes in ecDNA-positive samples.[40] Additionally, these findings are supported by evidence from other studies showing that inhibition of DNA repair pathways using a DNA-PKCS inhibitor can significantly reduce the frequency of ecDNA formation.[41] Interestingly, tumors classified as having a dominant ecDNA subtype showed the lowest HRD scores, suggesting that a relatively intact homologous recombination pathway may contribute to their comparatively better survival outcomes.

The identification of *PGC-1α* amplification via ecDNA in our dataset represents a rare but potentially significant event. While *PGC-1α* overexpression has been linked to GB, its genomic drivers remain unclear. A review of published datasets, including the Genomics England cohort ($n = 291$),[38] did not report *PGC-1α* ecDNA events, though other rare ecDNA amplifications (e.g., *MET*, *MYCN*, *PSIP1*) were noted, highlighting the heterogeneous nature of ecDNA in GB. In our case, the patient harbored *PGC-1α* ecDNA and showed the highest enrichment for both PPR and MTC pathway signatures, consistent with the gene's known role in mitochondrial biogenesis and oxidative phosphorylation. Although this represents a single case, it points to a possible ecDNA-driven mechanism of metabolic reprogramming in GBM. Larger studies with integrated ecDNA profiling are needed to determine whether *PGC-1α* amplification recurs in additional patients.

While extensive genomic studies on GB exist, they primarily focus on retrospective populations with The Cancer Genome Atlas (TCGA) and International Cancer Genome Consortium (ICGC). This study addresses this gap by providing a detailed characterization of ecDNA and other complex genomic alterations in a Hong Kong prospective patient cohort. Notably, patients with tumors enriched in MES-like, potentially driven by *NF1* mutations and *EGFR* ecDNA, exhibited worse survival outcomes. Given that ecDNA is increasingly recognized as the driver of therapy resistance, the data suggest a more carefully selected targeted therapy approach and offer opportunities for targeted therapy such as the S-phase checkpoint kinase CHK1 inhibitor, which has shown sensitivity toward ecDNA-amplified tumor cells.[42] These findings underscore the need for tailored therapeutic approaches based on distinct molecular subtypes.

### Limitations of the study

Our research revealed the role of ecDNA in driving tumor heterogeneity and reprogramming the TME. Nonetheless, several limitations should be acknowledged. Experimental validation is essential to establish the causal contribution of ecDNA to tumor evolution and immune modulation. *In vitro* perturbation assays and *in vivo* models that directly target ecDNA-associated vulnerabilities will be critical to demonstrate the mechanistic link and therapeutic relevance. Moreover, translating these insights into actionable clinical strategies remains a major challenge. Future studies incorporating molecular stratification and prospective patient recruitment will be required to enable the application of ecDNA biology toward improved patient management and treatment outcomes in glioblastoma.

## RESOURCE AVAILABILITY

### Lead contact

Requests for further information and resources should be directed to and will be fulfilled by the lead contact, Brian Chung (bhychung@hku.hk).

### Materials availability

This study did not generate new, unique reagents.

### Data and code availability

- This paper analyzes existing, publicly available snRNA-seq and scATAC-seq datasets, accessible at CellxGene (https://cellxgene.cziscience.com/collections/999f2a15-3d7e-440b-96ae-2c806799c08c) and the TCGA publication page (https://gdc.cancer.gov/about-data/publications/TCGA-ATAC-Seq-2024)[29]
- This paper analyzes existing, publicly available ChIP-seq datasets of H3K27Ac, accessible at GEO (GEO: GSE145646).[30]
- This paper analyzes existing, publicly available RNA-seq data, accessible at the CGGA database (http://www.cgga.org.cn/).
- Raw sequencing data derived from human samples have been deposited at the Hong Kong Genome Institute's Synergistic Research Environment (SRE). In accordance with the data-sharing policy of the Hong Kong Genome Project (HKGP), the raw sequencing data are made available to researchers upon request to the Hong Kong Genome Institute (hkgi_gc_team@genomics.org.hk). In addition, both the snRNA-seq raw count matrix for each sample and the corresponding metadata have been deposited in a secure AWS environment hosted in Hong Kong (https://hkgi-gbm-ecdna-study.s3.ap-east-1.amazonaws.com/GBM_cellranger_output.zip; https://hkgi-gbm-ecdna-study.s3.ap-east-1.amazonaws.com/Metadata.pdf) and are publicly available as of the date of publication.
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

## ACKNOWLEDGMENTS

The HKGP is a publicly funded genome sequencing initiative commissioned by the Health Bureau of the HKSAR Government. The authors would like to thank Dr. Su-vui Lo, Chief Executive Officer of the Hong Kong Genome Institute, for instrumental leadership and guidance. We thank all the laboratory staff at HKGI for their help with sample collection, processing, and sequencing. We also thank the Department of Neurosurgery Brain Tumor Biobank and the neurosurgeons at HKU and Queen Mary Hospital for their support and contributions.

## AUTHOR CONTRIBUTIONS

Conceptualization, B.H.Y.C., A.E.H., A.T.W.C., and W.T.; methodology, W.T., W.L.L., and C.W.S.L.; investigation, W.T., W.L.L., C.W.S.L., W.M., A.H.Y.T., D. Y., and J.S.L.K.; writing – original draft, W.T.; writing – review & editing, B.H.Y. C., A.E.H., A.T.W.C., G.K.K.L., and D.S.; resources, K.M.-Y.K. and G.K.K.L.; supervision, B.H.Y.C. and A.E.H.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

## STAR★METHODS

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

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
    - ○ Short-read WGS sample processing and sequencing

  - ○ Long-read WGS sample processing and sequencing
  - ○ Single-nucleus RNA sequencing sample processing and sequencing
  - ○ Short-read WGS data processing and analysis
  - ○ ecDNA detection and characterization
  - ○ Long-read WGS data processing and analysis
  - ○ Single nucleus transcriptomic analysis
  - ○ hdWGCNA analysis
  - ○ Cell abundance calculation in bulk RNA-seq data
  - ○ ATAC-seq signal comparison between ecDNA-positive and ecDNA-negative tumors
  - ○ Cellular dynamics analysis
  - ○ Cell-cell interaction analysis
  - ○ Survival analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS

## SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2025.116426.

Received: February 28, 2025
Revised: July 11, 2025
Accepted: September 23, 2025
Published: October 12, 2025

## REFERENCES

1. Woo, P.Y.M., Yau, S., Lam, T.-C., Pu, J.K.S., Li, L.-F., Lui, L.C.Y., Chan, D.T.M., Loong, H.H.F., Lee, M.W.Y., Yeung, R., et al. (2023). Patterns of care and survival of Chinese glioblastoma patients in the temozolomide era: a Hong Kong population-level analysis over a 14-year period. Neuro-Oncol. Pract. *10*, 50–61. https://doi.org/10.1093/nop/npac069.
2. Louis, D.N., Perry, A., Wesseling, P., Brat, D.J., Cree, I.A., Figarella-Branger, D., Hawkins, C., Ng, H.K., Pfister, S.M., Reifenberger, G., et al. (2021). The 2021 WHO Classification of Tumors of the Central Nervous System: a summary. Neuro Oncol. *23*, 1231–1251. https://doi.org/10.1093/neuonc/noab106.
3. Brennan, C.W., Verhaak, R.G.W., McKenna, A., Campos, B., Noushmehr, H., Salama, S.R., Zheng, S., Chakravarty, D., Sanborn, J.Z., Berman, S.H., et al. (2013). The Somatic Genomic Landscape of Glioblastoma. Cell *155*, 462–477. https://doi.org/10.1016/j.cell.2013.09.034.
4. Sosinsky, A., Ambrose, J., Cross, W., Turnbull, C., Henderson, S., Jones, L., Hamblin, A., Arumugam, P., Chan, G., Chubb, D., et al. (2024). Insights for precision oncology from the integration of genomic and clinical data of 13,880 tumors from the 100,000 Genomes Cancer Programme. Nat. Med. *30*, 279–289. https://doi.org/10.1038/s41591-023-02682-0.
5. Kinnersley, B., Sud, A., Everall, A., Cornish, A.J., Chubb, D., Culliford, R., Gruber, A.J., Lärkeryd, A., Mitsopoulos, C., Wedge, D., and Houlston, R. (2024). Analysis of 10,478 cancer genomes identifies candidate driver genes and opportunities for precision oncology. Nat. Genet. *56*, 1868–1877. https://doi.org/10.1038/s41588-024-01785-9.
6. Yang, L., Jia, R., Ge, T., Ge, S., Zhuang, A., Chai, P., and Fan, X. (2022). Extrachromosomal circular DNA: biogenesis, structure, functions and diseases. Signal Transduct. Target. Ther. *7*, 342. https://doi.org/10.1038/s41392-022-01176-8.
7. Yan, X., Mischel, P., and Chang, H. (2024). Extrachromosomal DNA in cancer. Nat. Rev. Cancer *24*, 261–273. https://doi.org/10.1038/s41568-024-00669-8.
8. Chu, A.T.W., Tong, A.H.Y., Tse, D.M.S., Lo, C.W.S., Lau, C.C.F., Li, C.Y.F., Tai, N.S.Y., Wong, L.W., Choy, G.K.C., Tse, B.Y.Y., et al. (2023). The Hong Kong genome project: building genome sequencing capacity and capability for advancing genomic science in Hong Kong. J. Transl. Genet. Genom. *7*, 196–212. https://doi.org/10.20517/jtgg.2023.22.
9. Ruiz-Moreno, C., Salas, S.M., Samuelsson, E., Minaeva, M., Ibarra, I., Grillo, M., Brandner, S., Roy, A., Forsberg-Nilsson, K., Kranendonk, M.E.G., et al. (2025). Charting the Single-Cell and Spatial Landscape of IDH-Wildtype Glioblastoma with GBmap. Neuro Oncol. noaf113. https://doi.org/10.1093/neuonc/noaf113.
10. Garofano, L., Migliozzi, S., Oh, Y.T., D'Angelo, F., Najac, R.D., Ko, A., Frangaj, B., Caruso, F.P., Yu, K., Yuan, J., et al. (2021). Pathway-based classification of glioblastoma uncovers a mitochondrial subtype with therapeutic vulnerabilities. Nat. Cancer *2*, 141–156. https://doi.org/10.1038/s43018-020-00159-4.
11. Yerly, L., Andreatta, M., Garnica, J., Nardin, C., Domizio, J.D., Aubin, F., Gilliet, M., Carmona, S.J., and Kuonen, F. (2025). Wounding triggers invasive progression in human basal cell carcinoma. Preprint at bioRxiv. https://doi.org/10.1101/2024.05.31.596823.
12. Zhao, Z., Zhang, K.-N., Wang, Q., Li, G., Zeng, F., Zhang, Y., Wu, F., Chai, R., Wang, Z., Zhang, C., et al. (2021). Chinese Glioma Genome Atlas (CGGA): A Comprehensive Resource with Functional Genomic Data from Chinese Glioma Patients. Genom., Proteom. Bioinform. *19*, 1–12. https://doi.org/10.1016/j.gpb.2020.10.005.
13. Newman, A.M., Steen, C.B., Liu, C.L., Gentles, A.J., Chaudhuri, A.A., Scherer, F., Khodadoust, M.S., Esfahani, M.S., Luca, B.A., Steiner, D., et al. (2019). Determining cell type abundance and expression from bulk tissues with digital cytometry. Nat. Biotechnol. *37*, 773–782. https://doi.org/10.1038/s41587-019-0114-2.
14. Chu, T., Wang, Z., Pe'er, D., and Danko, C.G. (2022). Cell type and gene expression deconvolution with BayesPrism enables Bayesian integrative analysis across bulk and single-cell RNA sequencing in oncology. Nat. Cancer *3*, 505–517. https://doi.org/10.1038/s43018-022-00356-3.
15. Deshpande, V., Luebeck, J., Nguyen, N.-P.D., Bakhtiari, M., Turner, K.M., Schwab, R., Carter, H., Mischel, P.S., and Bafna, V. (2019). Exploring the landscape of focal amplifications in cancer using AmpliconArchitect. Nat. Commun. *10*, 392. https://doi.org/10.1038/s41467-018-08200-y.
16. Zhu, K., Jones, M.G., Luebeck, J., Bu, X., Yi, H., Hung, K.L., Wong, I.T.-L., Zhang, S., Mischel, P.S., Chang, H.Y., and Bafna, V. (2024). CoRAL accurately resolves extrachromosomal DNA genome structures with long-read sequencing. Genome Res. *34*, 1344–1354. https://doi.org/10.1101/gr.279131.124.
17. Ng, A.W.T., McClurg, D.P., Wesley, B., Zamani, S.A., Black, E., Miremadi, A., Giger, O., Hoopen, R.T., Devonshire, G., Redmond, A.M., et al. (2024). Disentangling oncogenic amplicons in esophageal adenocarcinoma. Nat. Commun. *15*, 4074. https://doi.org/10.1038/s41467-024-47619-4.
18. Kolmogorov, M., Yuan, J., Lin, Y., and Pevzner, P.A. (2019). Assembly of long, error-prone reads using repeat graphs. Nat. Biotechnol. *37*, 540–546. https://doi.org/10.1038/s41587-019-0072-8.
19. Tan, Z., Luo, X., Xiao, L., Tang, M., Bode, A.M., Dong, Z., and Cao, Y. (2016). The Role of PGC1α in Cancer Metabolism and its Therapeutic Implications. Mol. Cancer Ther. *15*, 774–782. https://doi.org/10.1158/1535-7163.mct-15-0621.
20. Bruns, I., Sauer, B., Burger, M.C., Eriksson, J., Hofmann, U., Braun, Y., Harter, P.N., Luger, A.-L., Ronellenfitsch, M.W., Steinbach, J.P., and Rieger, J. (2019). Disruption of peroxisome proliferator–activated receptor γ coactivator (PGC)-1α reverts key features of the neoplastic phenotype of glioma cells. J. Biol. Chem. *294*, 3037–3050. https://doi.org/10.1074/jbc.ra118.006993.
21. Conte, M., Matteuzzi, T., Esposito, A., Chiariello, A.M., Bianco, S., Vercellone, F., and Nicodemi, M. (2023). Phase separation of ecDNA aggregates establishes in-trans contact domains boosting selective MYC regulatory interactions. Preprint at bioRxiv. https://doi.org/10.1101/2023.07.17.549291.
22. Handa, H., Honma, K., Oda, T., Kobayashi, N., Kuroda, Y., Kimura-Masuda, K., Watanabe, S., Ishihara, R., Murakami, Y., Masuda, Y., et al. (2020). Long Noncoding RNA PVT1 Is Regulated by Bromodomain Protein BRD4 in Multiple Myeloma and Is Associated with Disease Progression. Int. J. Mol. Sci. *21*, 7121. https://doi.org/10.3390/ijms21197121.
23. Mermel, C.H., Schumacher, S.E., Hill, B., Meyerson, M.L., Beroukhim, R., and Getz, G. (2011). GISTIC2.0 facilitates sensitive and confident

localization of the targets of focal somatic copy-number alteration in human cancers. Genome Biol. *12*, R41. https://doi.org/10.1186/gb-2011-12-4-r41.

24. Negrini, S., Gorgoulis, V.G., and Halazonetis, T.D. (2010). Genomic instability — an evolving hallmark of cancer. Nat. Rev. Mol. Cell Biol. *11*, 220–228. https://doi.org/10.1038/nrm2858.
25. Ben-David, U., and Amon, A. (2020). Context is everything: aneuploidy in cancer. Nat. Rev. Genet. *21*, 44–62. https://doi.org/10.1038/s41576-019-0171-x.
26. Knijnenburg, T.A., Wang, L., Zimmermann, M.T., Chambwe, N., Gao, G.F., Cherniack, A.D., Fan, H., Shen, H., Way, G.P., Greene, C.S., et al. (2018). Genomic and Molecular Landscape of DNA Damage Repair Deficiency across The Cancer Genome Atlas. Cell Rep. *23*, 239–254.e6. https://doi.org/10.1016/j.celrep.2018.03.076.
27. Singh, V.K., Rastogi, A., Hu, X., Wang, Y., and De, S. (2020). Mutational signature SBS8 predominantly arises due to late replication errors in cancer. Commun. Biol. *3*, 421. https://doi.org/10.1038/s42003-020-01119-5.
28. Touat, M., Li, Y.Y., Boynton, A.N., Spurr, L.F., Iorgulescu, J.B., Bohrson, C.L., Cortes-Ciriano, I., Birzu, C., Geduldig, J.E., Pelton, K., et al. (2020). Mechanisms and therapeutic implications of hypermutation in gliomas. Nature *580*, 517–523. https://doi.org/10.1038/s41586-020-2209-9.
29. Sundaram, L., Kumar, A., Zatzman, M., Salcedo, A., Ravindra, N., Shams, S., Louie, B.H., Bagdatli, S.T., Myers, M.A., Sarmashghi, S., et al. (2024). Single-cell chromatin accessibility reveals malignant regulatory programs in primary human cancers. Science *385*, eadk9217. https://doi.org/10.1126/science.adk9217.
30. Xu, L., Chen, Y., Huang, Y., Sandanaraj, E., Yu, J.S., Lin, R.Y.-T., Dakle, P., Ke, X.-Y., Chong, Y.K., Koh, L., et al. (2021). Topography of transcriptionally active chromatin in glioblastoma. Sci. Adv. *7*, eabd4676. https://doi.org/10.1126/sciadv.abd4676.
31. Weiler, P., Lange, M., Klein, M., Pe'er, D., and Theis, F. (2024). CellRank 2: unified fate mapping in multiview single-cell data. Nat. Methods *21*, 1196–1205. https://doi.org/10.1038/s41592-024-02303-9.
32. Jackson, C., Cherry, C., Bom, S., Dykema, A.G., Wang, R., Thompson, E., Zhang, M., Li, R., Ji, Z., Hou, W., et al. (2025). Distinct myeloid-derived suppressor cell populations in human glioblastoma. Science *387*, eabm5214. https://doi.org/10.1126/science.abm5214.
33. Governa, V., de Oliveira, K.G., Bång-Rudenstam, A., Offer, S., Cerezo-Magaña, M., Li, J., Beyer, S., Johansson, M.C., Månsson, A.-S., Edvardsson, C., et al. (2024). Protumoral lipid droplet–loaded macrophages are enriched in human glioblastoma and can be therapeutically targeted. Sci. Transl. Med. *16*, eadk1168. https://doi.org/10.1126/scitranslmed.adk1168.
34. Efremova, M., Vento-Tormo, M., Teichmann, S.A., and Vento-Tormo, R. (2020). CellPhoneDB: inferring cell–cell communication from combined expression of multi-subunit ligand–receptor complexes. Nat. Protoc. *15*, 1484–1506. https://doi.org/10.1038/s41596-020-0292-x.
35. Chen, Y., Huo, R., Kang, W., Liu, Y., Zhao, Z., Fu, W., Ma, R., Zhang, X., Tang, J., Zhu, Z., et al. (2023). Tumor-associated monocytes promote mesenchymal transformation through EGFR signaling in glioma. Cell Rep. Med. *4*, 101177. https://doi.org/10.1016/j.xcrm.2023.101177.
36. Khan, F., Pang, L., Dunterman, M., Lesniak, M.S., Heimberger, A.B., and Chen, P. (2023). Macrophages and microglia in glioblastoma: heterogeneity, plasticity, and therapy. J. Clin. Investig. *133*, e163446. https://doi.org/10.1172/jci163446.
37. Tang, W., Lo, C.W.S., Ma, W., Chu, A.T.W., Tong, A.H.Y., and Chung, B.H.Y. (2024). Revealing the role of SPP1+ macrophages in glioma prognosis and therapeutic targeting by investigating tumor-associated macrophage landscape in grade 2 and 3 gliomas. Cell Biosci. *14*, 37. https://doi.org/10.1186/s13578-024-01218-4.
38. Bailey, C., Pich, O., Thol, K., Watkins, T.B.K., Luebeck, J., Rowan, A., Stavrou, G., Weiser, N.E., Dameracharla, B., Bentham, R., et al. (2024). Origins and impact of extrachromosomal DNA. Nature *635*, 193–200. https://doi.org/10.1038/s41586-024-08107-3.
39. Rahman, R., Shi, D.D., Reitman, Z.J., Hamerlik, P., de Groot, J.F., Haas-Kogan, D.A., D'Andrea, A.D., Sulman, E.P., Tanner, K., Agar, N.Y.R., et al. (2024). DNA damage response in brain tumors: A Society for Neuro-Oncology consensus review on mechanisms and translational efforts in neuro-oncology. Neuro Oncol. *26*, 1367–1387. https://doi.org/10.1093/neuonc/noae072.
40. Lin, M.S., Jo, S.-Y., Luebeck, J., Chang, H.Y., Wu, S., Mischel, P.S., and Bafna, V. (2024). Transcriptional immune suppression and up-regulation of double-stranded DNA damage and repair repertoires in ecDNA-containing tumors. eLife *12*, RP88895. https://doi.org/10.7554/elife.88895.
41. Dharanipragada, P., Zhang, X., Liu, S., Lomeli, S.H., Hong, A., Wang, Y., Yang, Z., Lo, K.Z., Vega-Crespo, A., Ribas, A., et al. (2023). Blocking Genomic Instability Prevents Acquired Resistance to MAPK Inhibitor Therapy in Melanoma. Cancer Discov. *13*, 880–909. https://doi.org/10.1158/2159-8290.cd-22-0787.
42. Tang, J., Weiser, N.E., Wang, G., Chowdhry, S., Curtis, E.J., Zhao, Y., Wong, I.T.-L., Marinov, G.K., Li, R., Hanoian, P., et al. (2024). Enhancing transcription–replication conflict targets ecDNA-positive cancers. Nature *635*, 210–218. https://doi.org/10.1038/s41586-024-07802-5.
43. Korotkevich, G., Sukhov, V., Budin, N., Shpak, B., Artyomov, M.N., and Sergushichev, A. (2021). Fast gene set enrichment analysis. Preprint at bioRxiv. https://doi.org/10.1101/060012.
44. Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. https://link.springer.com/book/10.1007/978-0-387-98141-3.
45. Tickle, T., Tirosh, I., Georgescu, C., Brown, M., and Haas, B. (2019). inferCNV of the Trinity CTAT Project. https://github.com/broadinstitute/inferCNV.
46. Islam, S.M.A., Díaz-Gay, M., Wu, Y., Barnes, M., Vangara, R., Bergstrom, E.N., He, Y., Vella, M., Wang, J., Teague, J.W., et al. (2022). Uncovering novel mutational signatures by de novo extraction with SigProfilerExtractor. Cell Genom *2*, 100179. https://doi.org/10.1016/j.xgen.2022.100179.
47. Mayakonda, A., Lin, D.-C., Assenov, Y., Plass, C., and Koeffler, H.P. (2018). Maftools: efficient and comprehensive analysis of somatic variants in cancer. Genome Res. *28*, 1747–1756. https://doi.org/10.1101/gr.239244.118.
48. Luebeck, J., Huang, E., Kim, F., Liefeld, T., Dameracharla, B., Ahuja, R., Schreyer, D., Prasad, G., Adamaszek, M., Kenkre, R., et al. (2024). AmpliconSuite: an end-to-end workflow for analyzing focal amplifications in cancer genomes. Preprint at bioRxiv. https://doi.org/10.1101/2024.05.06.592768.
49. Talevich, E., Shain, A.H., Botton, T., and Bastian, B.C. (2016). CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing. PLoS Comput. Biol. *12*, e1004873. https://doi.org/10.1371/journal.pcbi.1004873.
50. Li, H. (2021). New strategies to improve minimap2 alignment accuracy. Bioinformatics *37*, 4572–4574. https://doi.org/10.1093/bioinformatics/btab705.
51. Wick, R.R., Schultz, M.B., Zobel, J., and Holt, K.E. (2015). Bandage: interactive visualization of de novo genome assemblies. Bioinformatics *31*, 3350–3352. https://doi.org/10.1093/bioinformatics/btv383.
52. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. Cell *184*, 3573–3587.e29. https://doi.org/10.1016/j.cell.2021.04.048.
53. Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., Baglaenko, Y., Brenner, M., Loh, P.R., and Raychaudhuri, S. (2019). Fast, sensitive and accurate integration of single-cell data with Harmony. Nat. Methods *16*, 1289–1296. https://doi.org/10.1038/s41592-019-0619-0.
54. Morabito, S., Reese, F., Rahimzadeh, N., Miyoshi, E., and Swarup, V. (2023). hdWGCNA identifies co-expression networks in high-dimensional transcriptomics data. Cell Rep. Methods *3*, 100498. https://doi.org/10.1016/j.crmeth.2023.100498.
55. La Manno, G., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V., Lidschreiber, K., Kastriti, M.E., Lönnerberg, P., Furlan, A., et al.

(2018). RNA velocity of single cells. Nature *560*, 494–498. https://doi.org/10.1038/s41586-018-0414-6.

56. Bergen, V., Lange, M., Peidli, S., Wolf, F.A., and Theis, F.J. (2020). Generalizing RNA velocity to transient cell states through dynamical modeling. Nat. Biotechnol. *38*, 1408–1414. https://doi.org/10.1038/s41587-020-0591-3.
57. Troulé, K., Petryszak, R., Cakir, B., Cranley, J., Harasty, A., Prete, M., Tuong, Z.K., Teichmann, S.A., Garcia-Alonso, L., and Vento-Tormo, R. (2025). CellPhoneDB v5: inferring cell–cell communication from single-cell multiomics data. Nat. Protoc., 1–29. https://doi.org/10.1038/s41596-024-01137-1.
58. Therneau, T.M., and Grambsch, P.M. (2000). Modeling Survival Data: Extending the Cox Model. https://CRAN.R-project.org/package=survival.
59. Biecek, A.K., Marcin Kosinski, P. (2025). survminer: Drawing Survival Curves using “ggplot2.” https://github.com/kassambara/survminer.
60. Granja, J.M., Corces, M.R., Pierce, S.E., Bagdatli, S.T., Choudhry, H., Chang, H.Y., and Greenleaf, W.J. (2021). ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. Nat. Genet. *53*, 403–411. https://doi.org/10.1038/s41588-021-00790-6.
61. Behera, S., Catreux, S., Rossi, M., Truong, S., Huang, Z., Ruehle, M., Visvanath, A., Parnaby, G., Roddey, C., Onuchic, V., et al. (2024). Comprehensive and accurate genome analysis at scale using DRAGEN accelerated algorithms. Preprint at bioRxiv. https://doi.org/10.1101/2024.01.02.573821.
62. Olson, N.D., Wagner, J., McDaniel, J., Stephens, S.H., Westreich, S.T., Prasanna, A.G., Johanson, E., Boja, E., Maier, E.J., Serang, O., et al. (2022). PrecisionFDA Truth Challenge V2: Calling variants from short and long reads in difficult-to-map regions. Cell Genom. *2*, 100129. https://doi.org/10.1016/j.xgen.2022.100129.
63. Investigators, T.A. of U.R.P.G., Group, M.W., Bick, A.G., Metcalf, G.A., Mayo, K.R., Lichtenstein, L., Rura, S., Carroll, R.J., Musick, A., Linder, J. E., et al. (2024). Genomic data in the All of Us Research Program. Nature *627*, 340–346. https://doi.org/10.1038/s41586-023-06957-x.
64. Scheffler, K., Catreux, S., O’Connell, T., Jo, H., Jain, V., Heyns, T., Yuan, J., Murray, L., Han, J., and Mehio, R. (2023). Somatic small-variant calling methods in Illumina DRAGEN$^{TM}$ Secondary Analysis. Preprint at bioRxiv. https://doi.org/10.1101/2023.03.23.534011.
65. Telli, M.L., Timms, K.M., Reid, J., Hennessy, B., Mills, G.B., Jensen, K.C., Szallasi, Z., Barry, W.T., Winer, E.P., Tung, N.M., et al. (2016). Homologous Recombination Deficiency (HRD) Score Predicts Response to Platinum-Containing Neoadjuvant Chemotherapy in Patients with Triple-Negative Breast Cancer. Clin. Cancer Res. *22*, 3764–3773. https://doi.org/10.1158/1078-0432.ccr-15-2477.
66. Gu, Z., Eils, R., and Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics *32*, 2847–2849. https://doi.org/10.1093/bioinformatics/btw313.
67. Chakravarty, D., Gao, J., Phillips, S.M., Kundra, R., Zhang, H., Wang, J., Rudolph, J.E., Yaeger, R., Soumerai, T., Nissan, M.H., et al. (2017). OncoKB: A Precision Oncology Knowledge Base. JCO Precis. Oncol. *2017*, 1–16. https://doi.org/10.1200/po.17.00011.
68. Zeng, A.G.X., Bansal, S., Jin, L., Mitchell, A., Chen, W.C., Abbas, H.A., Chan-Seng-Yue, M., Voisin, V., van Galen, P., Tierens, A., et al. (2022). A cellular hierarchy framework for understanding heterogeneity and predicting drug response in acute myeloid leukemia. Nat. Med. *28*, 1212–1223. https://doi.org/10.1038/s41591-022-01819-x.
69. van Dijk, D., Sharma, R., Nainys, J., Yim, K., Kathail, P., Carr, A.J., Burdziak, C., Moon, K.R., Chaffer, C.L., Pattabiraman, D., et al. (2018). Recovering Gene Interactions from Single-Cell Data Using Data Diffusion. Cell *174*, 716–729.e27. https://doi.org/10.1016/j.cell.2018.05.061.

## STAR★METHODS

### KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| Biological samples | | |
| GB patients' tumor tissue and blood | This study | N/A |
| Critical commercial assays | | |
| QIAsymphony SP DNA Midi Kit | Qiagen | Cat# 937255 |
| KAPA HyperPlus kit PCR-free | Roche | Cat# 07962436001 |
| KAPA Unique Dual-Indexed Adapter kit | Roche | Cat# 08861919702 |
| KAPA HyperPure Beads | Roche | Cat# 08963860001 |
| NovaSeq 6000 S4 Reagent kit v1.5 (300 cycles) | Illumina | Cat# 20028312 |
| MagAttract HMW DNA Kit | Qiagen | Cat# 67563 |
| Ligation Sequencing Kit | Oxford Nanopore Technologies | Cat# SQK-LSK114 |
| PromethION Flow Cell R10 | Oxford Nanopore Technologies | Cat# FLO-PRO114M |
| Nuclei Extraction Buffer | Miltenyi Biotec | Cat# 130-128-024 |
| Anti-Nucleus MicroBeads | Miltenyi Biotec | Cat# 130-132-997 |
| Chromium Next GEM Single Cell 3′ Kit v3.1 | 10x Genomics | Cat# PN-1000268 |
| Chromium Next GEM Chip G Single Cell Kit | 10x Genomics | Cat# PN-1000120 |
| Dual Index Kit TT Set A | 10x Genomics | Cat# PN-1000215 |
| Deposited data | | |
| snRNA-seq data | This study | https://hkgi-gbm-ecdna-study.s3.ap-east-1.amazonaws.com/GBM_cellranger_output.zip<br>https://hkgi-gbm-ecdna-study.s3.ap-east-1.amazonaws.com/Metadata.pdf |
| Reference scRNA-seq data: GBMap | Moreno et al.[9] | https://cellxgene.cziscience.com/collections/999f2a15-3d7e-440b-96ae-2c806799c08c |
| scATAC-seq | Sundaram et al.[29] | https://gdc.cancer.gov/about-data/publications/TCGA-ATAC-Seq-2024 |
| H3K27Ac ChIP-seq | Xu et al.[30] | GEO: GSE145646 |
| CGGA bulk RNA-seq data | Zhao et al.[12] | http://www.cgga.org.cn/ |
| Software and algorithms | | |
| DRAGEN (v4.0.3) | Illumina | https://sapac.illumina.com/products/by-type/informatics-products/dragen-secondary-analysis.html |
| CellRanger (v7.1.0) | 10x Genomics | http://www.10xgenomics.com |
| fgsea (v1.20.0) | Korotkevich et al.[43] | https://github.com/alserglab/fgsea |
| ggplot2 (v3.4.2) | Wickham et al.[44] | https://ggplot2.tidyverse.org/ |
| infercnvpy (v0.4.3) | Tickle et al.[45] | https://github.com/broadinstitute/inferCNV |
| SigProfilerExtractor | Islam et al.[46] | https://github.com/AlexandrovLab/SigProfilerExtractor |
| Maftools (v2.16.0) | Mayakonda et al.[47] | https://bioconductor.org/packages/devel/bioc/vignettes/maftools/inst/doc/maftools.html |
| R (v4.1.3) | R Core Team | https://www.r-project.org/ |
| GISTIC2.0 | Mermel et al.[23] | https://broadinstitute.github.io/gistic2/ |
| Ampliconsuite (v1.2.2) | Luebeck et al.[48] | https://github.com/AmpliconSuite |
| CNVkit (v0.9.10) | Talevich et al.[49] | https://cnvkit.readthedocs.io/ |
| Minimap2 (v2.24-r1122) | Li et al.[50] | https://github.com/lh3/minimap2 |
| wf-human-variation (v1.8.3) | EPI2ME | https://github.com/epi2me-labs/wf-human-variation |
| CoRAL | Zhu et al.[16] | https://github.com/AmpliconSuite/CoRAL |
| Flye (v2.9.3) | Kolmogorov et al.[18] | https://github.com/mikolmogorov/Flye |

(*Continued on next page*)

***Continued***

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| Bandange_ubuntu (v0.8.1) | Wick et al.[51] | https://github.com/rrwick/Bandage |
| Seurat (v4.3.0) | Hao et al.[52] | https://satijalab.org/seurat/ |
| Harmony (v0.1) | Korsunsky et al.[53] | https://github.com/immunogenomics/harmony |
| hdWGCNA (v0.3.3) | Morabito et al.[54] | https://smorabit.github.io/hdWGCNA/ |
| CIBERSORTx | Newman et al.[13] | https://cibersortx.stanford.edu/ |
| BayesPrism | Chu et al.[14] | https://github.com/Danko-Lab/BayesPrism |
| velocyto.py (v0.17.17) | Manno et al.[55] | https://velocyto2.org/ |
| Scvelo (v0.2.5) | Bergen et al.[56] | https://scvelo.readthedocs.io/ |
| CellRank (v2.0.6) | Weiler et al.[31] | https://cellrank.readthedocs.io/ |
| CellphonedDB (v5.0.1) | Troulé et al.[57] | https://github.com/ventolab/CellphoneDB |
| ktplotspy (v0.2.4) | Troulé et al.[57] | https://ktplotspy.readthedocs.io/ |
| survival (v3.5.3) | Terry et al.[58] | https://cran.r-project.org/web/packages/survival/index.html |
| survminer (v0.4.9) | Kassambara et al.[59] | https://cran.r-project.org/web/packages/survminer/index.html |
| ArchR (v1.0.2) | Granja et al.[60] | https://www.archrproject.com/ |

## EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Frozen surgical specimens from 42 IDH wild-type GB patients (age 22–81 years; 26 male, 16 female; all self-identified as Asian) were included in this study. Paired blood samples from 41 patients were collected as germline controls. All patients provided written informed consent, and the study was approved by the Institutional Ethics Review Board of the University of Hong Kong and Hong Kong West Cluster (UW 07–273). Detailed clinical data, including patient demographics, treatment history, tumor characteristics, and survival outcomes, were collected from the electronic health record system. Due to the limited sample size, sex- and gender-based analyses were not performed.

## METHOD DETAILS

### Short-read WGS sample processing and sequencing

Samples were prepared and sequenced at the Hong Kong Genome Institute.[8] High-quality genomic DNA was extracted from fresh frozen tumor tissues and paired blood samples using the QIAsymphony SP system and QIAsymphony SP DNA Midi Kit (Qiagen). PCR-free WGS libraries were prepared using the KAPA HyperPlus kit and the KAPA Unique Dual-Indexed Adapter kit (Roche), following the manufacturer's protocol. Briefly, 1 μg of gDNA was enzymatically fragmented at 37°C for 15 min, end-repaired, 3′dA-tailed, ligated to dual-index adapters, and size-selected. The libraries were quantified by quantitative PCR using the KAPA Library Quantification Kit (Roche) and then sequenced on an Illumina NovaSeq 6000 platform to reach an average coverage of 100x for tumor samples and 30x for blood samples. The median insert size was 525 bp.

### Long-read WGS sample processing and sequencing

High-molecular-weight genomic DNA was extracted from frozen tumor tissues using the MagAttract HMW DNA Kit (Qiagen). Library preparation was performed using a Ligation Sequencing Kit (SQK-LSK114, Oxford Nanopore Technologies, ONT) according to the manufacturer's protocol. Sequencing was performed using PromethION Flow Cell R10 (M Version) (FLO-PRO114M) on the ONT PromethION platform.

### Single-nucleus RNA sequencing sample processing and sequencing

Approximately 25 mg of fresh-frozen tumor tissue was transferred from dry ice into pre-chilled gentleMACS C Tubes containing 1.5 mL of ice-cold Nuclei Extraction Buffer supplemented with 0.2 U/μL RNase inhibitor. Fresh frozen tumor samples were dissociated using gentleMACS Dissociator. Nuclei were isolated using Anti-Nucleus MicroBeads and magnetic separation with LS columns according to the manufacturer's protocol. The isolated nuclei were counted, and loaded onto the Chromium Controller X (10x Genomics) using Chromium Chip G to capture approximately 5,000–10,000 nuclei per sample. Nuclei were encapsulated in gel beads to form gel bead-in-emulsions (GEMs). After reverse transcription within the GEMs, the emulsions were broken, and the barcoded complementary DNA (cDNA) was purified and amplified. The cDNA was fragmented, A-tailed, ligated with adapters, and indexed. The resulting libraries were sequenced on an Illumina NovaSeq 6000 platform with a minimum sequencing depth of 50,000 reads per nucleus.

### Short-read WGS data processing and analysis

The Illumina Dragen platform (v4.0.3) was utilized for base-calling and variant calling, which ranked top as the best-performing short-read call set in the precision FDA Truth Challenge and is widely used in large genome projects.[4,61–64] Briefly, reads were demultiplexed, mapped, and aligned to the human reference genome (GRCh38), followed by SNV, SV, and CNV calling. The DRAGEN pipeline produced sample-level metrics, such as mapping quality, ploidy estimation, SV, and CNV metrics.

The HRD score was computed using the Illumina DRAGEN Somatic pipeline (v4.0.3). This workflow uses tumor and matched normal BAM files to generate copy number calls. Based on these, the following three genomic instability metrics were calculated: LOH; TAI; LST. The total HRD score is defined as the sum of these three metrics, consistent with the original definition by Telli et al.[65]

SigProfilerExtractor[46] was used to analyze the mutational signatures and extract the SBS, DBS, and ID signatures. One hundred iterations were performed for the SigProfiler signature extraction. A signature was considered true if it was similar to a COSMIC V3.4 signature (cosine similarity $\geq$0.90) or if it could be reconstructed using multiple COSMIC V3.4 signatures with a reconstruction cosine similarity $\geq$0.90. Maftools (2.16.0)[47] and ComplexHeatmap (v2.16.0)[66] were used to visualize SNVs. GISTIC2.0 was used to identify focal gain and loss regions.[23]

### ecDNA detection and characterization

Ampliconsuite (v1.2.2)[15] was employed to detect ecDNA. Initially, regions with copy numbers greater than four and sizes exceeding 10kbp were defined as regions of interest (ROIs) (CNVkit, v0.9.10). Breakpoints within these ROIs were identified using a combination of CNV and SV analysis. AmpliconArchitect then automatically searches for other intervals participating in the amplicon and performs breakpoint graph construction. This process partitions all intervals into segments and builds an amplicon graph, optimizing a balanced flow on the graph to account for the copy numbers in each segment. Subsequently, AmpliconClassifier (v1.1.2)[48] was used to categorize the amplicons into different types, including circular amplicons, breakage–fusion–bridge amplifications, heavily rearranged amplifications, and linear amplifications, as previously described.[67]

### Long-read WGS data processing and analysis

For long-read data, whole-genome sequences were mapped to the reference genome hg38 using Minimap2 (v2.24-r1122). Structural variant calling and modified base calling were carried out using Sniffles2 (v2.0.7) and Modkit that implemented in the human variation workflow (v1.8.3) from Nanopore.

CoRAL[16] was utilized to analyze ecDNA structures with long-read sequencing data. Initially, segmenting intervals at positions with an increase in copy number (CN > 6) were identified. Subsequently, a breakpoint graph was constructed, and cycle decomposition was performed on the amplified seeds. To conduct *de novo* assembly, we extracted reads mapping to the amplified regions and employed the long-read assembly algorithm Flye (v2.9.3-b1797)[18] to create an assembly map for each amplicon, as previously described.[17] Subsequently, we annotated each graph using the genes identified in the amplicons. The assembled graph was visualized using Bandange_ubuntu (v0.8.1).[51] Dominant ecDNA is defined as a single ecDNA species constituting more than 80% of the total ecDNA population within a sample. Low count is defined as a total ecDNA copy number less than 11. Co-existence refers to the presence of multiple ecDNA species within a sample, with no single species exceeding 80% of the total ecDNA population.

### Single nucleus transcriptomic analysis

snRNA-seq data were obtained by aligning the reads to the human genome (GRCh38 Ensemble: v98) using CellRanger v7.1.0, with a median of 90,462 mean reads per cell. Nuclei with fewer than 200 detected genes, more than 100,000 UMIs, or over 15% mitochondrial genes were filtered out. Dimension reduction and unsupervised clustering were performed according to the standard workflow in Seurat (v4.3.0).[52] Harmony (v0.1) was used to perform batch-effect correction. The snRNA-seq data were then mapped onto a publicly available GB atlas (GBMap)[9] by identifying the anchors and integrating the data. Large-scale copy number alterations were inferred using inferCNVpy (v0.4.3), with tumor cells defined by a CNV score over 0.02.

### hdWGCNA analysis

Weighted Gene Co-expression Network Analysis (hdWGCNA) was conducted utilizing the hdWGCNA (v0.3.3).[54] The presence of *EGFR* ecDNA, copy number of *EGFR*, and methylation score of *EGFR* intron 1 [chr7: 55109000–55112000] were considered as three traits. The optimal soft power was determined as 13. Dimensionality reduction and visualization of the co-expression network were achieved using the ModuleUMAPPlot function. Subsequently, Gene Ontology analysis was performed on the top 100 genes associated with the module, employing the compareCluster function. The module scores were projected onto a UMAP using the FeaturePlot function.

### Cell abundance calculation in bulk RNA-seq data

To infer cell type composition from bulk RNA-seq profiles, we employed two independent deconvolution methods: CIBERSORTx[13] and BayesPrism,[14] using custom single-nucleus-derived reference matrices constructed from our GBM snRNA-seq dataset. We generated a reference matrix using gene expression counts from 17 annotated clusters in the snRNA-seq dataset. To ensure robust and specific gene signatures, we filtered genes expressed in at least 200 cells within each cluster. The minimum expression threshold for inclusion in the reference matrix was set to 0.25. CIBERSORTx deconvolution was then performed on transcript-per-million (TPM) normalized CGGA GBM bulk RNA-seq data using "S-mode" for batch correction and "absolute" mode for estimating absolute cell type fractions, as previously described.[37,68]

The CGGA TPM data also underwent deconvolution using the BayesPrism algorithm. Initially, the model was set up with the new.prism function. Subsequently, deconvolution was performed using the run.prism function. The final cell type fractions were derived from the posterior estimates utilizing the get.fraction function.

### ATAC-seq signal comparison between ecDNA-positive and ecDNA-negative tumors

ATAC-seq signal comparison between ecDNA-positive and ecDNA-negative tumors was conducted using scATAC-seq data processed with the ArchR framework (v1.0.2).[60] An ArchRProject was constructed, and doublets were filtered out. Sample-specific metadata, including ecDNA status and cell type annotations, were incorporated using addCellColData(). Samples were stratified based on copy number: $\leq 6$ ($n = 6$) and $>6$ ($n = 3$). To assess chromatin accessibility differences between ecDNA-positive and ecDNA-negative tumors, cells were grouped by ecDNA status, and pseudo-bulk replicates were constructed using addGroupCoverages(). Peak calling was performed using addReproduciblePeakSet() for each ecDNA group. The accessibility of *EGFR* and peak regions was visualized with plotBrowserTrack(), stratified by ecDNA status.

### Cellular dynamics analysis

RNA velocity analysis of the snRNA-seq dataset was performed using 'velocyto.py' (v0.17.17)[55] and 'scvelo' (v0.2.5).[56] BAM files generated by the 10x CellRanger pipeline were sorted and processed using 'velocyto run10x' to generate loom file, which were subsequently merged across *EGFR* ecDNA positive and negative patient groups. Cellular dynamics were analyzed using 'CellRank' (v2.0.6),[31] with Markov Affinity-based Graph Imputation of Cells (MAGIC) applied via 'magic-impute' (v3.0.0)[69] to enhance transcript smoothing. For each patient group (*EGFR* ecDNA-positive and -negative), a Velocity Kernel was computed separately. A Generalized Perron Cluster Cluster Analysis (GPCCA) estimator was employed to infer macrostates. The number of macrostates was determined using CellRank's Schur decomposition method. Fate probabilities were estimated by aggregating over all random walks, utilizing the 'direct' solver with the "ilu" preconditioner and use_petsc = True. Circular projections were used to visualize transitions toward different terminal states. This analysis generated a transition matrix representing the probability of each cell transitioning to another. Macrostates with the highest incoming or self-transition probabilities were designated as terminal macrostates. Gene expression dynamics along pseudotime were visualized using heatmaps and trajectory-based gene trend analysis.

### Cell-cell interaction analysis

CellphonedDB (v5.0.1)[57] was used to analyze cell-cell interactions between TAM and malignant cells. Cell barcodes and corresponding cell type annotations were extracted from the single-cell dataset and formatted as metadata for the analysis. To identify significantly enriched ligand-receptor pairs, we applied cpdb_statistical_analysis_method.call, which performs permutation-based statistical analysis to detect significantly enriched interactions within the given cell type. For visualization, ktplotspy (v0.2.4) was used to generate interaction networks, with a specific focus on the "Signaling by Epidermal Growth Factor" pathway.

### Survival analysis

The patients' overall survival across different levels of MES-like cells and gene signature scores among various patient subgroups, was evaluated using Mantel-Cox Log Rank tests with the 'survival' (v3.5.3). Survival curves were visualized using Kaplan–Meier plots with the 'survminer' (v0.4.9).

## QUANTIFICATION AND STATISTICAL ANALYSIS

Data were analyzed using the R software (v4.1.3) for all the statistical analyses. Kaplan–Meier analysis with log rank test was used to determine survival differences between the groups. Statistical comparisons between two groups were evaluated using two-tailed Wilcoxon rank-sum test. Correlation analysis was performed using Spearman's or Pearson's correlation coefficient test. Significance was defined as $p$ value <0.05.