

# A proteogenomic atlas of 1032 brain metastases identifies molecular subtypes, immune landscapes, and therapeutic vulnerabilities

Received: 10 September 2025

Accepted: 15 January 2026

Published online: 26 January 2026

 Check for updates

Zhenyu Yang<sup>1,12</sup>, Shiyu Wei<sup>1,12</sup>, Hao Duan<sup>2,12</sup>, Xiuqi Wang<sup>3,12</sup>, Dainan Zhang<sup>4,5,12</sup>, Karrie Mei-Yee Kiang<sup>6,7,12</sup>, Yulan Deng<sup>1,12</sup>, Yuanzhong Yang<sup>8,12</sup>, Yin Ku<sup>1</sup>, Wanming Hu<sup>8</sup>, Chaowei Zou<sup>9</sup>, Meiqin Tang<sup>2</sup>, Yu Jiang<sup>10</sup>, Yu Liu<sup>3</sup>, Hongbin Lan<sup>3</sup>, Zexin Chen<sup>11</sup>, Gilberto Ka-Kit Leung<sup>6,7</sup>✉, Wang Jia<sup>4,5</sup>✉, Yonggao Mou<sup>2</sup>✉, Lunxu Liu<sup>1</sup>✉ & Gao Zhang<sup>3</sup>✉

The molecular heterogeneity of brain metastases hampers therapeutic development for cures. To address this unmet and urgent need, we construct a comprehensive multi-omic, single cell, and spatially resolved atlas of 1,032 pan-cancer brain metastases, identifying four robust molecular subtypes with distinct biological programs and clinical associations. These brain metastases subtypes (BrMS) are defined by unique biological states: neural-like (BrMS1), metabolic (BrMS3), highly proliferative/immune-excluded (BrMS4), and an immune-infiltrated (BrMS2) state featuring a coordinated epithelial-mesenchymal transition program. Patient-derived organoids coupled with targeted drug screening indicate subtype-specific molecular dependencies and putative targets, notably mTOR signaling activation in BrMS3 and CDK4/6 axis activation in BrMS4, while BrMS1 and BrMS2 display distinct radiobiologic and immunologic signatures. This atlas provides a rigorous classification framework of BrMs and offers insights into subtype-specific molecular vulnerabilities.

Brain metastasis (BrM) is one of the most frequent complications in patients with tumors. Approximately 30% of patients with solid tumors develop BrMs, and the incidence has markedly increased in the past 20 years given the longer duration of survival and better prognosis with progress made in the treatment landscape of malignant solid tumors<sup>1,2</sup>. Lung cancer, breast cancer, melanoma, and colorectal cancer contribute most to BrMs, which account for 60–80% of patients<sup>2</sup>. Current approaches to the treatment of BrMs include local therapy (such as whole-brain radiotherapy, stereotactic radiosurgery and surgical resection) and systemic therapy (such as chemotherapy, targeted therapy and immunotherapy)<sup>1,2</sup>.

Despite these advancements, standard of care therapies for BrMs are still limited.

Recent advances in molecular profiling technologies have provided unprecedented insights into the genomic, transcriptomic and epigenomic landscapes of BrMs, which provides potentially druggable targets to treat BrMs. Brastianos et al. characterized the genomic landscape of 86 pan-cancer BrMs, and found that clinically actionable alterations present in about 53% of patients with BrMs which were not detected in primary tumors, highlighting the importance of molecular characterization of metastatic lesions<sup>3</sup>. Shih et al. analyzed whole-exome sequencing (WES) data derived from 73 patients with BrMs of

A full list of affiliations appears at the end of the paper. ✉ e-mail: [gilberto@hku.hk](mailto:gilberto@hku.hk); [noanswear@hotmail.com](mailto:noanswear@hotmail.com); [mouyg@sysucc.org.cn](mailto:mouyg@sysucc.org.cn); [lunxu\\_liu@aliyun.com](mailto:lunxu_liu@aliyun.com); [gzhang6@me.com](mailto:gzhang6@me.com)

lung adenocarcinoma, and revealed that amplification frequencies of *MYC*, *YAPI* and *MMP13* were increased in BrMs, and overexpression of *MYC*, *YAPI* or *MMP13* promoted the incidence of BrMs<sup>4</sup>. In addition, Fischer et al. performed RNA sequencing (RNA-seq) on 88 BrMs of melanoma and 42 patient-matched extracranial metastases, and identified a significant level of immunosuppression and an enrichment of oxidative phosphorylation (OXPHOS) in BrMs of melanoma compared to patient-matched extracranial metastases<sup>5</sup>. Recent advances in metastatic tumor research, particularly the comprehensive MSKCC analysis of over 25,000 patients, have demonstrated that genomic alterations in metastases often exhibit cancer type-specific patterns and generally display greater chromosomal instability and clonal selection compared to primary tumors<sup>6</sup>. Our recent study integrated genomic, transcriptomic, proteomic and metabolomic data which were derived from 154 patients with matched and unmatched primary lung cancer and BrMs and suggested that mitochondrial-specific metabolism was activated but tumor immune microenvironment (TiME) was suppressed in BrM of lung cancer. Further, we demonstrated that the combination of a specific mitochondrial inhibitor gamitrinib and an anti-PD-1 immunotherapy significantly improved survival of mice bearing BrM of lung cancer<sup>7</sup>. Fukumura et al. analyzed transcriptomic and reverse-phase protein array data from 35 patients with matched primary or extracranial metastasis and BrMs from breast cancer, lung cancer and renal cell cancer, and showed similar findings<sup>8</sup>.

Tumor classification has emerged as a powerful approach to understanding tumor heterogeneity and guiding personalized treatment strategies. Molecular classification of tumors has revolutionized cancer management in various malignancies, such as breast cancer (luminal A, luminal B, HER2-enriched, and basal-like subtypes)<sup>9</sup>, glioblastoma (GBM) (proneural, neural, classical, and mesenchymal subtypes)<sup>10</sup>, colorectal cancer (microsatellite instability immune, canonical, metabolic, and mesenchymal subtypes)<sup>11</sup>. These molecular subtypes often correlate with distinct clinical outcomes and treatment responses, enabling more precise prognostication and leading to tailored therapeutic approaches. Previous studies have attempted to manage BrMs according to molecular subtypes of their primary tumors, and substantial improvements in survival have been achieved in patients with molecular subgroups whose alterations can be targeted with specific molecular compounds<sup>1</sup>. BrMs progressing from *EGFR*-mutant and *ALK*-rearranged non-small-cell lung cancer (NSCLC)<sup>12,13</sup>, *HER2*-positive breast cancer<sup>14</sup> and *BRAF*-mutant melanoma<sup>15</sup> can be successfully targeted with specific inhibitors. And immune-checkpoint inhibitors (ipilimumab, nivolumab, pembrolizumab and atezolizumab) have clearly improved the outcome of patients with BrMs from melanoma, NSCLC and triple-negative breast cancer, particularly in patients with asymptomatic brain metastases<sup>16–23</sup>.

Besides, while foundational studies have mapped the genomic alterations in BrMs, we lack a unifying framework that connects these alterations to the unique tumor microenvironment (TME) of the brain. It remains unclear how the interplay between tumor cells and the neural niche dictates distinct biological states, immune landscapes, and ultimately, therapeutic vulnerabilities. Crucially, most studies have lacked the spatial resolution to observe these interactions in situ and the functional models to test their clinical relevance. Here, we test the hypothesis that BrMs, irrespective of their primary origin, converge on a limited number of molecular subtypes defined by distinct immune landscapes and metabolic programs.

In this work, we present the cohort of 1032 pan-cancer BrMs derived from diverse primary tumor sites, along with 82 matched primary tumors and 20 GBMs, which integrates both publicly available data and newly generated samples (Supplementary Data 1–2). For this newly generated cohort, we extensively profile them using RNA-seq, WES, 4D data-independent acquisition (4D-DIA) proteomics, targeted metabolomics, and integrate diverse cutting-edge approaches such as

immunohistochemical staining (IHC), single-nucleus RNA sequencing (snRNA-seq)<sup>24</sup>, spatial transcriptomics (ST) with 10x Genomics Visium, spatial proteomics with PhenoCycler-Fusion 2.0, multiplex immunofluorescence (mIF) technologies and patient-derived organoid (PDO) models. Our findings reveal four reproducible BrMs subtypes with distinct immune and metabolic programs and highlight subtype-specific molecular targets, establishing a TME based framework for understanding BrM biology.

## Results

### Patient cohort and consensus clustering identifies four distinct subtypes of pan-cancer BrMs with differential survival outcomes

To establish a comprehensive landscape of pan-cancer BrMs at the genomic, transcriptomic and proteomic levels, we retrospectively collected fresh frozen (FF) tumor tissues from Sun Yat-sen University Cancer Center (SYSUCC) (Guangzhou, China) ( $n = 59$ ) and Queen Mary Hospital (QMH) (Pok Fu Lam, Hong Kong, China) ( $n = 48$ ) (Fig. 1a and Supplementary Fig. 1a). Additionally, to contrast with primary brain tumors, we retrospectively collected 20 FF tumor tissues of GBM from SYSUCC. Following quality control (QC), these tumor specimens were subject to bulk RNA-seq, encompassing 107 BrMs and 20 GBMs. A subset of these tumor samples also underwent WES (78 BrMs), 4D-DIA quantitative proteomics (107 BrMs and 20 GBMs) and targeted metabolomics (74 BrMs), respectively, due to tissue availability and QC. In order to provide additional insights on both single cell and spatial levels, we employed snRNA-seq, IHC, mIF, ST and spatial proteomics to further investigate a subset, representative pan-cancer BrM samples.

To further assemble a larger cohort of pan-cancer BrMs, we integrated our data with multi-omics data derived from 5 published studies, which were subsequently designated as the discovery cohort in this study<sup>5,7,8,25,26</sup>. In total, our discovery cohort of pan-cancer BrM consisted of 356 pan-cancer BrM tumor specimens available for the multi-omic analysis, alongside with 82 patient-matched primary tumors (3 newly generated and 79 from published studies) and 20 GBMs. The demographic and clinical-pathological characteristics of these patients are detailed in Table 1. The most prevalent primary tumor sites for this discovery cohort were lung (34.6%), melanoma (29.2%), breast (17.7%), and kidney (4.8%), with other origins such as colon (3.7%) comprising the remainder (Fig. 1b). A comprehensive list of tumor origins is provided in Supplementary Data 1.

To validate our findings, we assembled an independent cohort consisting of 676 pan-cancer BrMs, which includes 227 newly generated, previously unpublished RNA-seq datasets (In-house Validation,  $n = 207$ ; BrM-PDO,  $n = 20$ ) from three major medical centers in China (Beijing Tiantan Hospital [BJTTH] [Beijing, China], SYSUCC, and West China Hospital [WCH] [Chengdu, China]), and 449 cases from 12 published studies<sup>27–38</sup>. The published datasets were categorized based on their gene expression profiling platforms: RNA-seq (External Validation 1,  $n = 202$ , refs. 27–33), microarray (External Validation 2,  $n = 199$ , refs. 34–37) and targeted gene expression profiling (NanoString) (External Validation 3,  $n = 48$ , ref. 38). The BrM-PDO cohort was employed to generate PDO models for functional studies. The most prevalent primary tumor sites for this validation cohort were lung (45.8%), breast (24.7%), and melanoma (14.2%) (Supplementary Fig. 1b). The summary of patients' clinical information for this validation cohort is provided in Supplementary Data 2.

Taken together, we assembled the dataset of 1032 pan-cancer BrMs, accompanied by 82 matched primary tumors and 20 GBMs for the subsequent investigation on the molecular level. Sample numbers for each experimental assay, broken down by newly collected vs. publicly available data and by cohort (discovery vs. validation) are present in Table 2.

To define robust molecular subtypes at the transcriptional level, we performed batch correction to remove technical variation across cohorts using the ComBat-seq algorithm<sup>39</sup> (Supplementary Fig. 1c–f).



**Fig. 1 | Study workflow and molecular subtypes of pan-cancer BrMs reveal distinct biological features and clinical outcomes.**

**a** Study design overview. Created in BioRender. Wang, X. (2026) <https://BioRender.com/kjgbf33>. Discovery cohort included 107 BrMs, 20 GBMs, and 3 matched primary tumors from SYSUCC and QMH. Validation cohort comprised 207 BrMs from BJTTTH, SYSUCC, and WCH, 20 BrM-PDOs from SYSUCC and 449 BrMs from public data. Multi-platform profiling encompassed transcriptomics, WES, 4D-DIA proteomics, targeted metabolomics, snRNA-seq, ST, IHC, mIF and PhenoCycler-Fusion 2.0. **b** Primary tumor distribution in the discovery cohort ( $n = 356$ ), highlighting predominant origins: lung (34.6%), melanoma (29.2%), breast (17.7%), kidney (4.8%), and colon (3.7%). Created in BioRender. Wang, X. (2026) <https://BioRender.com/hnfxmdk>. **c** Consensus clustering identified four molecular subtypes (BrMS1,  $n = 98$ ; BrMS2,  $n = 59$ ; BrMS3,  $n = 94$ ; BrMS4,  $n = 105$ ). Heatmap shows differential gene expression patterns with patients' clinical information. **d** In-house Validation ( $n = 207$ )

confirmed consistent subtype expression patterns. **e** Clinical characteristic distribution across subtypes. Two-sided Chi-square test. **f-i** BrMS subtypes show significant postoperative survival differences in both discovery cohort ( $n = 78/34/73/75$ , 260 of 356 patients with survival data available;  $P = 0.0017$ , log-rank test) and In-house Validation cohort ( $n = 41/26/57/29$ , 153 of 207 patients with survival data available;  $P = 8.30E-05$ , log-rank test). BrMS4 demonstrates consistently worse survival compared with other subtypes. GSEA reveals subtype-specific pathway enrichment: neurological pathways in BrMS1 (**j**), interleukin signaling in BrMS2 (**k**), fatty acid metabolism in BrMS3 (**l**), and cell cycle/mitosis pathways in BrMS4 (**m**). NES and Benjamini-Hochberg (BH) adjusted  $P$ -values indicated. **n** Representative IHC images of subtype-specific markers, including GFAP, IFI16, ACSL5, and TOP2A. Three samples per subtype, each marker on separate slides. Scale bars: 50  $\mu\text{m}$ . Source data is provided as a Source Data file. NET, neuroendocrine tumor.

**Table 1 | Clinical characteristics of discovery cohorts**

Features	BrM from SYSUCC & QMH ( $n = 107$ )	Fukumura K et al. 2021 ( $n = 31$ )	GBM from SYSUCC ( $n = 20$ )	Duan Hao et al. 2024 ( $n = 46$ )	Fischer G et al. 2019 ( $n = 88$ )	Routh E et al. 2022 ( $n = 19$ )	Wardell C et al. 2021 ( $n = 65$ )
Sex							
Male, $n$ (%)	55 (51%)	10 (32%)	14 (70%)	30 (65%)	53 (60%)	0	33 (51%)
Female, $n$ (%)	52 (49%)	21 (68%)	6 (30%)	16 (35%)	35 (40%)	19 (100%)	32 (49%)
Primary site							
Lung	40 (37%)	10 (32%)	NA	46 (100%)	0	0	27 (42%)
Breast	21 (20%)	14 (45%)	NA	0	0	19 (100%)	9 (14%)
Kidney	8 (7%)	7 (23%)	NA	0	0	0	2 (3%)
Colon	10 (9%)	0	NA	0	0	0	3 (5%)
Melanoma	2 (2%)	0	NA	0	88 (100%)	0	14 (22%)
Other	26 (24%)	0	NA	0	0	0	10 (15%)
MGMT							
Methylated	NA	NA	7 (35%)	NA	NA	NA	NA
Unmethylated	NA	NA	9 (45%)	NA	NA	NA	NA
Unknown	NA	NA	4 (20%)	NA	NA	NA	NA
Age at craniotomy date (years mean (range))	58.3 (33.2, 75)	57.2 (25.6, 78.1)	47.1 (16.5, 73.3)	58.4 (29, 76.5)	52.4 (26, 76)	NA	NA
Post-operative survival (Craniotomy to censored in month) (months median, (range))	26.6 (0.6, 171.2) 5 missing	25.9 (0.3, 151)	12.1 (2.7, 25.5)	28.7 (2.5, 132) 7 missing	15.7 (0.6, 132.6)	NA	NA

Comprehensive summary of clinical and demographic characteristics for brain metastases (BrM) and glioblastoma multiforme (GBM) patient cohorts, including primary tumor origins and relevant clinical parameters.

Template Prediction (NTP) analysis<sup>41</sup> across three independent validation cohorts (Inhouse Validation, External Validation 1, External Validation 2), we observed most samples met the significance threshold (false discovery rate [FDR]  $< 0.01$ ) based on 1000 permutations, demonstrating robust classification consistency across different technological platforms (Supplementary Fig. 1o). The reliability of subtype assignments was further confirmed through silhouette score analysis, which showed strong clustering stability across all validation cohorts (Supplementary Fig. 1p).

We developed and validated a machine learning-based predictive model using a combined dataset from both discovery and validation cohorts (split as 70% training and 30% testing). This model demonstrated excellent performance with 97% accuracy on the training set and 84% accuracy on the test set. The model's robustness was further confirmed through cross-validation across multiple cohorts, consistently achieving high accuracy ( $> 0.875$ ) and adjusted Rand index (ARI) scores ( $> 0.575$ ) across validation cohorts. These comprehensive validation results support the reliability and reproducibility of our molecular classification system (Supplementary Fig. 1q). After performing leave-one-study-out cross-validation, we evaluated model

calibration, which showed well-calibrated probability estimates across studies (Supplementary Data 3).

The differential gene expression analysis revealed subtype-specific molecular signatures. BrMS1 was enriched with immune and neurological system-related genes, including *CD33*, *CX3CR1*, *CD163* and *IL10*, suggesting that M2-polarized macrophage and regulatory T cells infiltrated the TME. BrMS2 exhibited a higher expression level of genes related to immune response, inflammation (*IL-6*, *CCL2*, and *CXCL8*), and extracellular matrix (ECM) remodeling (*ITGB2*, *ITGAM*, *PLAU*, *MMP2*). BrMS3 exhibited an enrichment of epithelial markers (*EPCAM*, *KRT8*, *KRT18*, and *CDHI*) and metabolic genes (*NUDT8*, *AKR7A3* and *CYP3A5*), suggesting that a higher proportion of malignant epithelial cells in the TME of this subtype is characterized by the metabolic reprogramming. BrMS4 displayed an upregulation of cell cycle and proliferation-related genes (*CENPE*, *E2F1*, *MCM2*, *ORC1* and *CCNA2*). Moreover, BrMS4 exhibited the stem cell property with a higher expression level of *SOX2*, a neural stem cell marker, and *EZH2*, which maintains cellular pluripotency through H3K27 methylation (Fig. 1c). The complete differential expressed gene list is provided in Supplementary Data 4.

**Table 2 | Sample composition across experimental platforms and cohorts**

Assay Type	BrM Discovery Cohort	BrM Validation Cohort	Primary	GBM	Independent Cohort
WES					
Newly generated	78				
Public data	82				
Bulk Transcriptomic					
Newly generated	107	207	3	20	
Public data	249	449	79		
Bulk Proteomics (mass spectrometry)					
Newly generated	107		3	20	
Public data	11		11		
Targeted Metabolomics					
Newly generated	74				
Public data	12		12		
snRNA-seq					
Newly generated	16				
10x Genomics Visium					
Newly generated	12		4		
PhenoCycler-Fusion 2.0					
Newly generated	20	3			
DSP					
Public data					41 BrMs, 54 primary tumors
IHC					
Newly generated	12				
mIF					
Newly generated	6				
Public data	12				
Patient-derived Organoids (PDOs)					
Newly generated	14	4			20 BrM-PDOs

Comprehensive summary of sample numbers for each experimental assay, broken down by newly collected vs. publicly available data and by cohort (Discovery vs. Validation).

The analysis of patients' clinical characteristics revealed that while the distribution of primary tumor types varied significantly among subtypes, no substantial differences were observed in the distribution of sex or age (Fig. 1e). Specifically, BrMs originated from breast and colon tumors were primarily enriched in BrMS3. This enrichment of colon cancer-derived BrMs in BrMS3 remained significant after adjusting for study source (multinomial logistic regression,  $P < 0.05$ ) (Supplementary Data 5).

To evaluate the clinical relevance of our classification system, we performed multivariate Cox proportional hazards analysis, incorporating key clinical variables including primary tumor site, age, sex, and cohort. This analysis revealed BrMS4 as a significant independent prognostic factor (Supplementary Fig. 2a).

The survival analysis further demonstrated a significantly different prognosis, with BrMS2 showing a more favorable outcome and BrMS4 being associated with worse survival ( $P = 0.0017$ ) (Fig. 1f). When comparing BrMS4 against other subtypes, a significant survival disadvantage was observed ( $P = 0.00021$ ) (Fig. 1g). This distinct prognostic pattern was subsequently validated in three validation cohorts (Fig. 1h, i and Supplementary Fig. 2b–d).

To further investigate the molecular basis underlying four subtypes, we performed Gene Set Enrichment Analysis (GSEA) using Reactome and Hallmark gene sets from Molecular Signatures Database (MSigDB)<sup>42,43</sup>. This analysis revealed enrichment of subtype-specific pathways: neurological pathways in BrMS1, interleukin signaling in BrMS2, fatty acid metabolism in BrMS3, and cell cycle/mitosis-related pathways in BrMS4 (Supplementary Fig. 2e–f). To highlight the representative pathway features of each BrMS, we assessed the

similarity among enriched pathways and visualized the most relevant biological processes (Fig. 1j–m). These molecular characteristics were further validated at the protein level by performing the IHC staining of subtype-specific markers. We observed distinctive expression patterns of glial fibrillary acidic protein (GFAP) in BrMS1 (neural microenvironment), interferon gamma inducible protein 16 (IFI16) in BrMS2 (interleukin signaling), acyl-CoA synthetase long chain family member 5 (ACSL5) in BrMS3 (fatty acid metabolism), and DNA topoisomerase II alpha (TOP2A) in BrMS4 (cell cycle activation) (Fig. 1n).

Our further molecular characterization revealed that BrMS4 possessed distinct biological features. For instance, the transcriptional factor network analysis identified subtype-specific regulatory programs, with BrMS4 being characterized by proliferation-associated factors including *E2F7* and *FOXMI*, highlighting E2F as a central regulatory node (Supplementary Fig. 2g–i). Subsequently, the calculation of stemness score demonstrated stem cell properties were elevated in this subtype in comparison with BrMS1-3 (Supplementary Fig. 2j).

In summary, our analysis demonstrates that consensus molecular classification of 356 pan-cancer BrMs in the discovery cohort enables the identification of four molecularly distinct subtypes with unique biological identities: BrMS1 (neural-like), BrMS2 (immune-infiltrated), BrMS3 (metabolic), and BrMS4 (proliferative).

### The analysis of proteomics data reveals proteomic signatures for four subtypes of pan-cancer BrMs

To determine whether these transcriptomic subtypes are translated into distinct functional states at the protein level, we employed quantitative proteomics with 4D-DIA technology<sup>44,45</sup>, which integrates

retention time, mass-to-charge ratio, ion intensity, and ion mobility for reliable and reproducible protein quantification. In total, we quantified 9621 proteins across 118 pan-cancer BrMs in the discovery cohort including 107 newly generated and 11 from published studies (Supplementary Data 6). We performed differentially expressed proteins analysis and identified distinct proteomic profiles across four BrMS subtypes (Fig. 2a, Supplementary Data 7).

BrMS1 was enriched in glial and synaptic functions proteins (GFAP, NEFL and SYN1). BrMS2 was characterized by immune-related proteins, including HLA family members (HLA-DPB1, HLA-DRB1, HLA-DRA, HLA-DPA1, HLA-DMB and HLA-DRB5), immune cell markers (FCGR3A, CD74 and CD40), chemokines (CCL24), cathepsins (CTSL, CTSB, CTSS and CTSZ), and inflammatory mediators (EPX, PRG2 and PTGS1). BrMS3 showed upregulation of epithelial and metabolic proteins (EPCAM, CDH1, KRT7/15/18, GALNT3/6, ACSL5). BrMS4 exhibited increased expression of cell cycle and nucleic acid processing associated proteins (MCM family, TOP2A, STMN1, CDKN2A, TCF3, LSM5, SMARCC1).

Next, we performed integrative Reactome pathway analysis of both transcriptomic and proteomic data via ActivePathways<sup>46</sup> and revealed consistent biological distinctions with transcriptomic data (Fig. 2b–e, Supplementary Data 8). BrMS1 was dominated by neuronal signaling pathways such as neurotransmitter transport, synaptic signaling, RTK (Receptor tyrosine kinases)-MAPK (Mitogen-activated protein kinase) cascade, and NMDA (N-methyl-D-aspartic acid) receptor mediated synaptic plasticity. Stromal and immune pathways were enriched in BrMS2, including ECM organization, platelet activation, and immune system cascade. BrMS3 involved metabolic and stress response pathways, encompassing ER stress response, RNA processing and general metabolic regulation. BrMS4 exhibited enrichment of pathways related to proliferation, including cell cycle regulation, DNA repair and replication, transcriptional regulation, and viral response pathways. The correlation analysis of RNA transcripts and proteins correlation analysis showed that BrMS3 exhibited the highest mRNA-protein concordance among all subtypes with adjusting for tumor purity (Fig. 2f).

Beyond our analysis of canonical pathways, we applied Weighted Gene Co-expression Network Analysis (WGCNA)<sup>47</sup> to delineate functional protein modules and their hub components. This analysis identified 30 co-expression modules, 28 of which were annotated with specific biological functions (Fig. 2g, Supplementary Data 9). Mapping subtype-specific differentially expressed proteins onto these modules highlighted distinct module level expression patterns (Fig. 2h). Module eigengenes (ME) were calculated as the first principal components of each module's expression profile to capture overall co-expression trends (Fig. 2i). Of these, 25 modules exhibited significant differences among BrMS subtypes ( $P < 0.05$ ). Based on optimal ME values, samples were stratified into high and low expression groups for prognostic assessment (Fig. 2j).

Subsequently, we focused on four key modules (ME30, ME21, ME17, and ME15) displaying both subtype-specific expression and prognostic relevance: ME30 (OXPHOS) enriched in BrMS1, ME21 (EMT) in BrMS2, ME17 (Monocarboxylic acid catabolic process) in BrMS3, and ME15 (E2F targets) in BrMS4.

ME21, associated with an unfavorable prognosis, comprised ECM-related hub proteins (COL1A2, FBNI, COL4A2, FMOD, COL6A1, BGN) (Fig. 2k), reflecting roles in ECM organization, basement membrane assembly, and structural integrity. Patients with high ME21 expression showed significantly poorer survival compared with the ME21 Low group ( $P = 0.005$ ) (Fig. 2l).

Since neural-related modules, such as ME1 and ME6, were enriched in BrMS1 but showed no significant prognostic association. We next focused on ME30. This OXPHOS related module was modestly elevated in BrMS1 and lowest in BrMS4, suggestive of its widespread activity across BrMs rather than restriction to a specific subtype,

consistent with our previous findings. Additionally, ME30 was associated with favorable prognosis. Therefore, we further investigate its hub proteins to identify the potential regulators within this module.

Hub protein network of ME30 revealed MPV17L2, CKMT2, ATP6VOD2, PEX2, and MPC1 as core hub proteins, suggestive of their potential roles in mitochondrial energy homeostasis in BrMs (Supplementary Fig. 3a). ME17 is also related to metabolic functions, containing metabolic regulation and cell homeostasis regulators (MCCC2, GNL3L, APBB2, CNNM3, IDUA, SIGIRR), correlated with favorable prognosis (Supplementary Fig. 3b). ME15, characterized by DNA replication and cell cycle proteins (POLD1, MSH2/6, TOP2A, CDK4, and SMC2), is unfavorable in terms of prognosis (Supplementary Fig. 3c). The survival analysis further validated the association between prognosis and specific modules (Supplementary Fig. 3d–f). We performed a WGCNA analysis on the integrated proteomic data from 108 BrMs and 20 GBMs, and the modules obtained were highly similar to the results based on BrM data alone (Supplementary Fig. 3g–h).

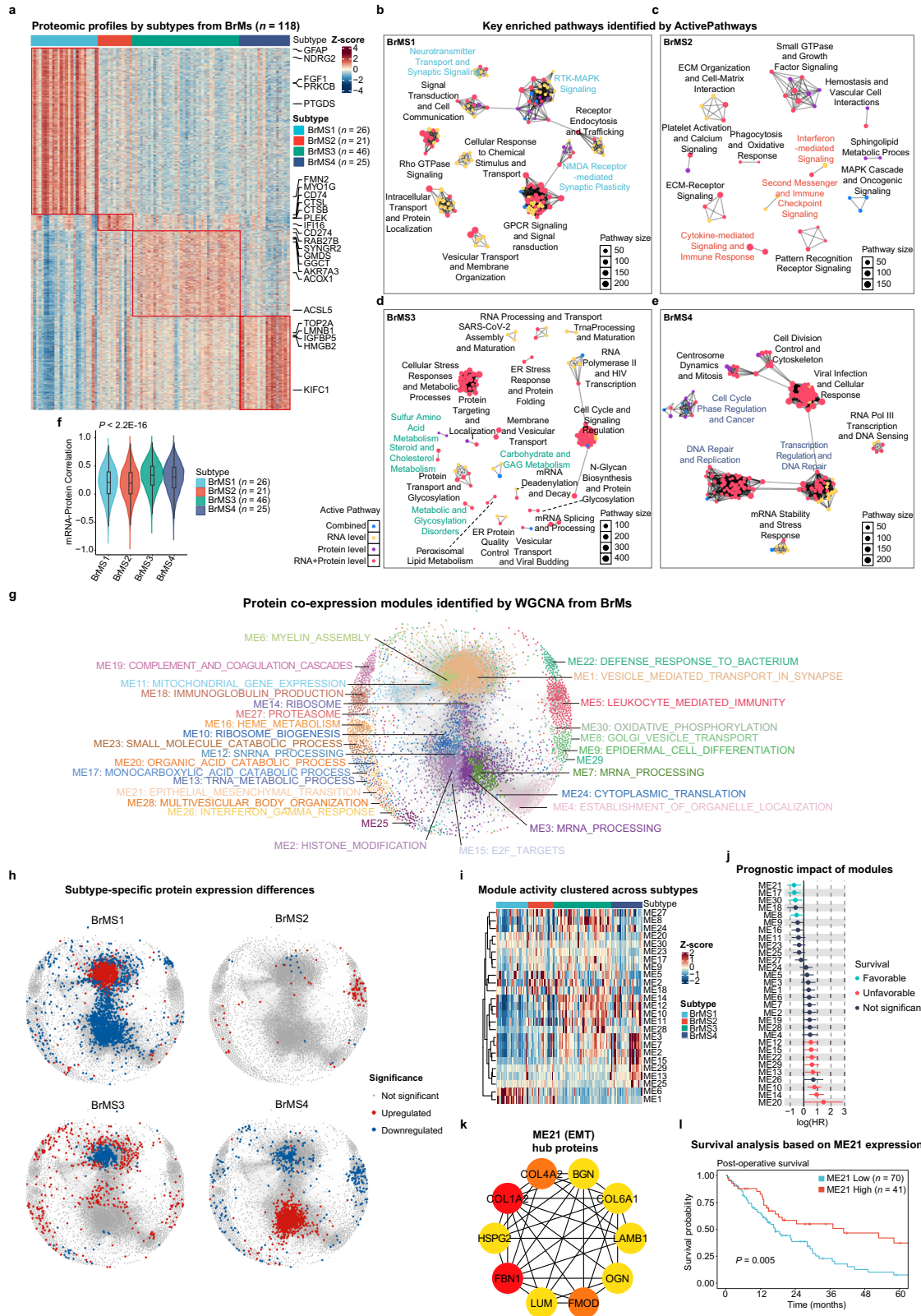
Together, our findings provide a proteomic level landscape of functional co-expression networks underlying four BrMS subtypes, highlighting distinct biological process and prognostically relevant modules.

### The genomic landscape of pan-cancer BrMs

Having defined the transcriptomic and proteomic landscapes, we next investigated whether these molecular subtypes are underpinned by distinct driver mutations or genomic instability patterns. The analysis of WES data was performed for comprehensive characterization of somatic mutations and copy number variations across samples. The genomic profiling of 160 BrMs (78 newly generated and 82 from published studies) revealed distinct mutational landscapes across four molecular subtypes (Supplementary Data 10). BrMS1 was characterized by mutations in key immune regulatory genes, including *TRAF2* and *CD248*. BrMS2 was dominated by alterations in the signal transduction pathway, exhibiting mutations in *IFNA10*, *CCER2*, *COL26A1* and *MORC4*. The metabolic subtype BrMS3 showed distinct mutations in *SLC26A5*, which was associated with metabolic reprogramming. The proliferative subtype BrMS4 was defined by genomic alterations in *PRAMEF15*, *MYH2*, *UACA* and especially *RBI*, dominantly affecting cell cycle control (Fig. 3a).

Across all subtypes, *TTN*, *TP53*, *MUC16* and *SYNE1* emerged as the most frequently altered genes, with mutational rates of 67%, 57%, 54%, and 48%, respectively (Supplementary Fig. 4a). These genes are also identified as significantly mutated in BrMs by previous studies<sup>7,48</sup>. The mutation pattern of *TP53* is consistent with previous reports showing frequent *TP53* alterations in BrMs from multiple cancer types including NSCLC and breast cancer<sup>49,50</sup>, indicating that the loss of its function may play a driving role in the development of BrMs.

Next, we performed the mutational signature analysis for each subtype based on purity ploidy corrected genomic profiles inferred by Sequenza<sup>51</sup>. Using non-negative matrix factorization, we identified four major mutational signatures, which were compared with known signatures in the Catalog of Somatic Mutations in Cancer (COSMIC v2) database<sup>52</sup> (Fig. 3b). Our results showed that Signature 1 (COSMIC\_1, spontaneous deamination of 5-methylcytosine) showed higher exposure in BrMS1. This age-related endogenous mutational process is consistent with the immune-enriched, hypermutated phenotype of BrMS1 tumors. Signature 3 (COSMIC\_4, tobacco smoking-associated) and Signature 4 (COSMIC\_13, APOBEC cytidine deaminase (C > G) transversions) showed higher exposure in BrMS4 (Fig. 3c). APOBEC-associated mutagenesis causes genomic instability and transcriptional alterations in cancers<sup>53,54</sup>. To exclude potential bias from primary tumor composition, we examined signature exposures in lung cancer BrMs, where signature patterns remained consistent with those in the overall cohort (Supplementary Fig 4b). We replicated the mutational signature analysis with SigProfiler using default settings, confirming



our COSMIC-based results with excellent concordance (cosine similarity = 0.967) (Supplementary Fig 4c).

Despite this distinctive genomic characteristic, both TMB and mutant-allele tumor heterogeneity (MATH) scores remained consistent across four BrM subtypes (Fig. 3d, e). BMS4 demonstrated the highest levels of somatic copy number alterations (SCNA) count and chromosomal instability (CIN) score, suggesting the highest genomic

instability. In contrast, BrMS1 showed relatively low SCNA and CIN scores, with significantly lowest tumor purity. Pathway perturbation scores showed that TGF- $\beta$  and ErbB signaling pathways exhibited the most pronounced differences, with significantly lower perturbation scores in BrMS4 (Fig. 3f-i).

The CNV analysis revealed amplifications of high frequency in several key chromosomal regions, including 1q21.2, 3q26.32, 8q24.11,

**Fig. 2 | Proteomic characterization of pan-cancer BrM subtypes.** **a** Differentially expressed proteins across BrMS subtypes ( $n = 26/21/46/25$ ). Selected subtype-specific protein markers are annotated on the right. **b–e** Integrated RNA-protein pathway analysis using ActivePathway for each BrMS subtype. Node colors indicate activation patterns: blue (mixed), yellow (RNA-specific), purple (protein-specific), and red (concordant RNA-protein activation). Cluster labels indicate representative subtype-enriched pathway groups. **f** RNA-protein correlations using partial Spearman correlation controlling for tumor purity (estimated by Sequenza) (matched RNA-protein sample sizes:  $n = 26/21/46/25$ ;  $P < 2.2E-16$ , Kruskal–Wallis test). The central line represents the median, and the upper and lower boundaries of the boxes indicate the third quartile (Q3) and the first quartile (Q1), respectively. The upper and lower whiskers extend to the most extreme data points within 1.5 times the interquartile range from the box edges. **g** WGCNA analysis identified 30 distinct protein co-expression modules (ME1–ME30), with 28 annotated to functional categories. Individual proteins are represented as nodes and colored according to their module assignment. **h** Subtype-specific protein expression differences

mapped onto the protein-protein interaction network ( $n = 6783$  proteins). **i** Heatmap showing the module eigengene scores of 25 modules across BrMS subtypes, with statistically significant differences identified by Kruskal–Wallis test (FDR-adjusted). **j** Forest plot showing the prognostic impact of protein modules on patient survival. Log hazard ratios (logHR, central dot) and 95% CI (horizontal line) were calculated from proteomic data of 111 BrM samples with survival data available using Cox proportional hazards regression. Modules are colored by prognostic association: favorable (cyan,  $HR < 1$ ,  $P < 0.05$ ), unfavorable (red,  $HR > 1$ ,  $P < 0.05$ ), or not significant (gray,  $P \geq 0.05$ ). **k** Hub protein network of ME21 module (EMT, BrMS2-enriched). Top 10 hub proteins were identified using the MCC score in cytoHubba (Cytoscape). Nodes are colored by MCC score, ranging from yellow (low) to red (high). **l** Kaplan–Meier analysis comparing survival of patients stratified by ME21 module expression based on optimal expression threshold (ME21 Low,  $n = 70$ ; ME21 High,  $n = 41$ ) ( $P = 0.005$ , log-rank). Source data is provided as a Source Data file.

12q15, 16p11.2, and 17q12 (Supplementary Fig. 4d). These amplifications, occurring in over 25% of 160 pan-cancer BrMs, encompass important oncogenes such as *MYC* (8q24), *PIK3CA* (3q26.32), *ERBB2* (*HER2*) (17q12), and *MDM2* (12q15), consistent with previous genomic studies of BrMs<sup>3,4,7</sup> (Supplementary Data 11). Concurrently, we identified several deletion regions of high frequency, including 1p36, 3p12.3, 8p23.1, 9p21.3, 10q23.31, 17p11.2, and 18q21.31 (Supplementary Fig. 4e), potentially affecting crucial tumor suppressor genes such as *CDKN2A* (9p21.3) and *PTEN* (10q23.31) (Supplementary Data 12).

The analysis of structural variants (SV) using Manta<sup>55</sup> identified 4494 SVs across all samples, comprising 3270 breakends (BND), 856 deletions (DEL), 308 duplications (DUP), and 60 insertions (INS). Those genes most frequently affected by SVs included *MUC19* (22 events), *AKRIC8P*, *C16orf52*, *FSTL4* (12 events each), *OR8G2P* (11 events), and *OR8G1* (10 events). Notably, the SV distribution varied significantly across molecular subtypes, with BrMS1 subtype exhibiting significantly lower SV counts compared to BrMS3 subtype ( $P < 0.001$ ) and BrMS4 subtype ( $P < 0.01$ ) (Supplementary Fig. 4f).

To further investigate the functional impact of these CNAs, we integrated CNA data with patient-matched transcriptomics and proteomics profiles (Supplementary Fig. 4g–h). We identified 1588 genes exhibiting significant cis-correlations between CNAs and mRNA expression, and 288 genes showing significant cis-correlations between CNAs and protein expression. Among them, 21 genes displayed consistent cis-correlations across CNAs, mRNA, and protein levels (FDR  $< 0.05$ ). Among the focal amplifications, several key genes, including *CHD1L*, *SF3B4*, *PDCD6*, *NUDCD1*, *NAXD*, *CARS2*, *ARHGEF7*, *TUBGCP3*, *CUL4A*, exhibited consistent changes across both RNA and protein levels. Similarly, focal deletions predominantly affected genes including *NDUFAF2*, *REEP5*, *ATG12*, *TMED7*, *GLUD1*, *PGAM1*, *EPG5*, *C18orf25*, *ELAC1*, *TXNL1*, *TIMM21*, *CTDPI* (Supplementary Data 13).

The comparison of significant CNV events across BrMS subtypes revealed distinct patterns: BrMS4 presented the most extensive genomic alterations, with significant amplification at 19p12 and deletion at 5q21.1. BrMS3 was characterized by more deletions at 6q27 and 2q11.1 (Fig. 3j, Supplementary Data 14).

Next, we systematically analyzed genetic alterations across key oncogenic pathways, including MAPK pathway, AKT pathway, and pathways related to immune evasion, immune response, cell cycle and epithelial mesenchymal transition (EMT), and epigenetic regulation. Enrichment in activating mutations in *PIK3CA* and *EGFR* was described in previous finding<sup>3,56</sup>. The EGFR/MAPK axis demonstrates intricate regulatory complexity, with *EGFR* exhibiting high-frequency mutations in BrMS1 (37%) and consistent amplification across subtypes (15–27%), while *KRAS* mutation occurred variably, peaking in BrMS2 (20%). *PIK3CA* showed both mutation and amplification events across BrM subtypes (mutation rate: 10–29% and amplification rate: 29–71%) and downstream components of PI3K-AKT-mTOR pathway also presented

subtype-specific alterations. Both *AKT1* mutation and *mTOR* mutation were most frequent in BrMS1 (7% and 17%). Notably, *TP53* displayed consistently high mutational frequencies across all subtypes (50–76%), reaching 76% in BrMS4, suggestive of widespread defects in DNA damage repair and cell cycle checkpoints (Fig. 3k).

We conducted Multi-Omics Factor Analysis (MOFA)<sup>57</sup> to integrate analyses of our transcriptomic, proteomic, and genomic data. MOFA identified 10 latent factors, with Factor 1 demonstrating the strongest association with BrMS classification (Supplementary Fig. 5a–d). This factor exhibited significant differences among BrMS subtypes, particularly distinguishing BrMS1 from the other subtypes. The pathway analysis of Factor 1 revealed consistent patterns across both transcriptomic and proteomic levels. Specifically, we observed a strong activation of neuronal system-related pathways and synaptic transmission, alongside distinct metabolic features and molecular transport, which was positively correlated with Factor 1. On the other aspect, cell cycle and DNA repair pathways were negatively correlated with Factor 1 at the transcriptomic level, while RNA metabolism and epigenetic regulation negatively correlated with Factor 1 at proteomic level (Supplementary Fig. 5e–f). The integrative analysis further confirmed our previous findings.

Taken together, our data revealed distinct genomic landscapes across four BrM subtypes and highlight both common features and subtype-specific alterations that may inform therapeutic strategies, suggesting potential molecular subtype-specific vulnerabilities and prognostic indicators for clinical applications.

### Metabolic features and molecular vulnerability of BrMS

The enrichment for metabolic pathways in BrMS3 at the transcriptomic level prompted us to investigate whether this subtype exhibits a distinct metabolic state. Indeed, we undertook targeted metabolomic profiling of 74 BrM samples with fresh frozen tissues available to us to measure targeted 75 metabolites of energy metabolism (Supplementary Data 15). BrMS3 showed widespread upregulation of metabolites, particularly prominent in glycolytic intermediates (3-phosphoglycerate, 2-phosphoglycerate, and phosphoenolpyruvate), and nucleotide metabolism intermediates. BrMS1 and BrMS2 showed elevated levels of TCA cycle metabolites such as Acetyl-CoA, citric acid, isocitric acid, and cis-aconitic acid, while BrMS4 generally exhibited lower metabolic activity (Fig. 4a, b). The pathway analysis further highlighted BrMS3's distinct metabolic profile compared to other subtypes. The most significant differences were observed in the TCA cycle, glycolysis/gluconeogenesis pathway, and pentose phosphate pathway, with BrMS3 consistently showing higher pathway activities (Fig. 4c).

We performed an integrated analyses of RNA-seq and 4D-DIA proteomics data of 82 pan-cancer BrMs and matched primary tumors to identify molecular features specific to BrMs. The pathway



enrichment analysis revealed an enrichment of multiple metabolism-related pathways, such as OXPHOS and TCA cycle, indicating there was a substantial metabolic reprogramming in BrMs compared to primary tumors (Fig. 4d). Further analysis revealed distinct metabolic profiles between BrMS subtypes and their matched primary tumors at both the RNA and protein levels. (Supplementary Fig. 6a). While OXPHOS was consistently enriched across all subtypes, BrMS3 demonstrated the

most significant upregulation of citric acid cycle, steroid biosynthesis, and cholesterol biosynthesis pathways, suggestive of a unique metabolic reprogramming in this subtype. To validate these findings, we analyzed targeted metabolomics data which focused on energy metabolism, derived from 12 patient-matched pairs of primary lung cancers and BrMs, which revealed enhanced OXPHOS in BrMs compared to primary tumors<sup>7</sup>. The differential metabolite analysis revealed

**Fig. 3 | Genomic landscape of pan-cancer BrMs.** **a** Oncoplot showing top five significantly mutated genes in each subtype (BrMS1-4) from 160 BrM samples. Top panel shows clinical and molecular annotations including primary cancer type, sex, age, OS, and OS status. The bar plot above displays log-transformed mutation counts per sample. The main oncoplot below depicts the distribution of somatic mutations across samples (columns) and genes (rows). Different colors represent various mutation types. Samples are grouped by molecular subtypes (BrMS1-4). The sidebar shows the percentage of samples harboring mutations in each gene. **b** Mutational signature analysis identifying four major signatures and their correlation with COSMIC signatures. Stacked bar plot showing the relative contribution of four mutational signatures across BrMS1-4 subtypes, different colors indicate distinct mutational signatures, and each vertical bar represents an individual sample. One sample was excluded from signature analysis due to insufficient mutation counts for bootstrap estimation. **c** Mutational signature exposure rate across pan-cancer BrM subtypes ( $n = 159$ ). **d-i** Boxplots of log-scaled tumor

mutational burden (TMB) (**d**), mutant-allele tumor heterogeneity (MATH) score (**e**), tumor purity estimated using Sequenza (**f**), somatic copy number alteration (SCNA) counts (**g**), chromosomal instability (CIN) scores (**h**), and 11 key signaling pathway mutation perturbation scores (**i**) across BrMS1-4 subtypes ( $n = 41/30/55/34$ ) (two-sided Wilcoxon rank-sum test, exact  $P$ -values shown). In all boxplots, the central line represents the median, and the upper and lower boundaries of the boxes indicate the third quartile (Q3) and the first quartile (Q1), respectively. The upper and lower whiskers extend to the most extreme data points within 1.5 times the interquartile range from the box edges. **j** Pie charts comparing the frequency of four specific chromosomal deletions across subtypes, Fisher's test. **k** Genetic alteration frequencies of key genes in critical signaling pathways across BrMS subtypes. Each gene box displays four percentages representing the frequencies of genetic alterations (mutation, deletion, or amplification) in BrMS1-4 subtypes. Genes are grouped by signaling and functions. Source data is provided as a Source Data file.

an upregulation of key metabolic intermediates in BrMs (Fig. 4e, Supplementary Data 16), including glycolysis pathway metabolites such as glyceraldehyde-3-phosphate, dihydroxyacetone phosphate, and fructose-1,6-bisphosphate, and TCA cycle-related metabolites such as succinyl-CoA and acetyl-CoA. Additionally, the analysis revealed that multiple amino acids were significantly less abundant in BrMs. The pathway enrichment analysis further confirmed that glycolysis/gluconeogenesis and pentose phosphate pathway were among the most significantly enriched metabolic pathways, while lipoic acid metabolism pathways were most significantly increased (Fig. 4f).

Our previous study reported that targeting mitochondrial metabolism via gamitrinib showed promising efficacy in treatment of preclinical models of lung cancer BrMs<sup>7</sup>. To identify a subset of pan-cancer BrMs that would potentially benefit from gamitrinib, we employed a computational approach, VIPER-OncoTreat<sup>58,59</sup>. We first calculated the master regulators (MRs) in BrMs compared to primary tumors. Secondly, the inverting effects of gamitrinib on top 30 activated MRs and 30 inactivated MRs was estimated by 1-tailed analytic-rank based enrichment analysis (aREA) algorithm. Pan-cancer BrMs were predicted to be sensitive if normalized enrichment score (NES) < 0 and Bonferroni  $p_{adj} < 10^{-5}$ . The predictions were further validated using PDO drug screening (Fig. 4g). BrM-PDOs with  $IC_{50} < 2 \mu M$  were considered to be sensitive to gamitrinib (Supplementary Fig. 6b). For example, gamitrinib showed a high efficacy in one BrM-PDO (B282T), which resulted in reversed MR states (NES = -5.97,  $P < 0.0001$ ). Low  $IC_{50}$  value (0.44  $\mu M$ ) was confirmed in PDO validation experiment (Fig. 4h, i). In contrast, organoids derived from one BrM-PDO (B337T) showed resistance to gamitrinib, with forward MR states (NES = 3.54;  $P < 0.0001$ ). This phenotype of drug resistance was consistent with its high  $IC_{50}$  value (6.03  $\mu M$ ) (Fig. 4j, k). Finally, we performed sensitivity testing of gamitrinib in a total of 30 BrM-PDOs. Notably, the model achieved an accuracy of 0.8, with a sensitivity of 0.8696 and specificity of 0.5714 (Supplementary Fig 6c, Supplementary Data 17).

Given that radiotherapy is a standard of care for the treatment of BrMs<sup>60</sup>, we re-analyzed RNA-seq data from an independent cohort<sup>29</sup> with breast cancer BrMs from External Validation 1 to investigate the overall survival among four BrMS subtypes after radiotherapy. Our results showed that BrMS1 exhibited the most favorable survival association following radiotherapy. This association was still significant in multivariable analysis when adjusted for age, PAM50 subtype, and number of extracranial metastatic sites though (HR = 0.31, 95% CI: 0.13–0.75,  $P = 0.009$ ) (Fig. 4l, Supplementary Data 18).

Furthermore, we identified subtype-specific targets and pathway dependencies. The drug screening of 20 BrM-PDOs showed lower  $IC_{50}$  value for the mTOR inhibitor-everolimus in BrMS3 and for the CDK4/6 inhibitor-abemaciclib in BrMS4 compared with BrM-PDOs from other subtypes (Fig. 4m, n and Supplementary Fig. 6d–e). These results demonstrated enhanced sensitivity of BrMS3 to mTOR pathway

inhibition (e.g., everolimus) and selective vulnerability of BrMS4 to the CDK4/6 blockade (e.g., abemaciclib).

### The immune landscape of pan-cancer BrMs reveals a subtype of immunotherapy-responsive BrMs

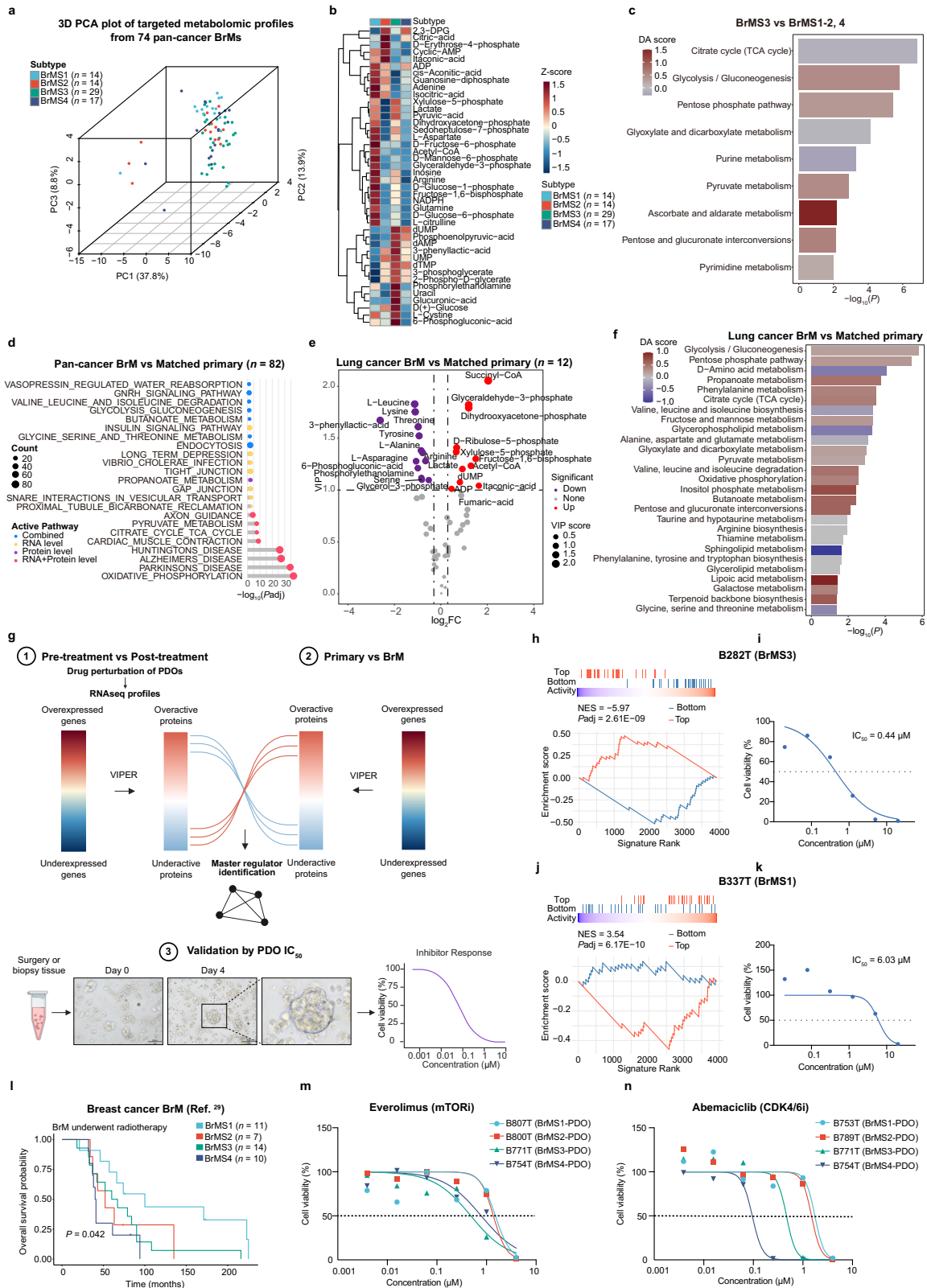
To further shed light on the TME of BrMs, we analyzed 20 GBM samples and 82 matched primary tumors alongside pan-cancer BrMs. Using the immune classification proposed by Thorsson et al.<sup>61</sup>, we grouped all BrMs, GBMs and primary tumors into six immune subtypes (C1–C6) (Supplementary Fig. 7a). C2 (IFN- $\gamma$  dominant) was the predominant subtype in both BrMs and primary tumors, whereas C4 (lymphocyte depleted) was markedly enriched in BrMs compared to primary tumors, indicating a shift towards a distinct TME.

Applying the microenvironment Cell Populations-counter (MCP-counter)<sup>62</sup>, we quantified infiltrating immune cells across four BrM subtypes in comparison with primary tumors and GBMs (Fig. 5a, Supplementary Data 19). Both BrMS1 and BrMS2 exhibited higher scores of cytotoxic T lymphocytes (CTLs) than BrMS3 and BrMS4 (Supplementary Fig. 7b). We also obtained consistent result with CIBERSORT<sup>63</sup> (Supplementary Fig. 7c).

The cytolytic index, which is calculated as the median expression of *GZMA* and *PRF1*<sup>64</sup>, was significantly higher in BrMS1 and BrMS2 than in BrMS3 and BrMS4 (Supplementary Fig. 7d), confirming the increase in cytotoxic activity. Stratification of BrMs from discovery cohort by CTL score showed that patients with more abundant CTL infiltration ( $n = 92$ ) had significantly better survival than those with less abundant CTL infiltration ( $n = 168$ ), indicating that CTL infiltration is a favorable prognostic indicator<sup>65</sup> (Fig. 5b). A high endothelial score also correlated with improved survival, consistent with previous report<sup>66</sup> (Supplementary Fig. 7e).

We next analyzed expression levels of MHC molecule and immune checkpoint genes across BrMS subtypes. BrMS2 showed the highest expression levels of both MHC class I and class II molecules, suggestive of an active antigen presentation (Fig. 5c–f). BrMS1 and BrMS2 exhibited higher expression of *PD-L1* and exhaustion markers (*CTLA4*, *HAVCR2*, *LAG3* and *TIGIT*) compared to BrMS3 and BrMS4 at both RNA and protein levels (Fig. 5g, h, Supplementary Fig. 7f). Furthermore, exhaustion markers expression positively correlated with cytotoxic lymphocyte infiltration levels (Supplementary Fig. 7g).

To further explore this relationship, we profiled subsets of CD8 T cells based on the classification of Chu et al.<sup>67</sup>. (Supplementary Fig. 7h). Scores of precursor exhausted (p-Tex, c7) CD8 T cells and exhausted (Tex, c1) CD8 T cells differed significantly among subtypes, being notably higher in BrMS1, BrMS2 and primary tumors (Supplementary Fig. 7i–j). Finally, by comparing pathway and cell type scores between BrMs and matched primary tumors, we found that elevated DNA damage response and repair pathways in BrMs were associated with poor prognosis, whereas increased CD8 T cell infiltration correlated with improved survival (Supplementary Fig. 7k).



To explore genomic determinants of immune infiltration, we identified amplification of chromosome 8p11.22 was significantly correlated with an increase in CTL levels (Supplementary Fig. 7i). Integrative pathway enrichment analysis (ActivePathways) combined differentially expressed genes and proteins upregulated in BrMs with 8p11.22 amplification, revealing significant enrichment in Valine Leucine and Isoleucine Degradation, Propanoate Metabolism and Fatty Acid

Metabolism (Supplementary Fig. 7m). Those metabolic pathways were associated with CTL infiltrating or exhausting according to previous reports<sup>68,69</sup>. In a multivariate Cox model adjusting for CTL score, primary tumor site, age, sex, study cohort, and TMB, the CTL score remained an independent prognostic factor, respectively (Supplementary Fig. 7n).

Building on the link between immune infiltration and cancer immunotherapy, we next examined the potential immunotherapeutic

**Fig. 4 | Metabolic features and therapeutic implications of BrMS subtypes.** **a** 3D PCA of targeted metabolomic profiles from 74 pan-cancer BrMs. **b** Heatmap of metabolite abundance across subtypes (BrMS1-4,  $n = 14/14/29/17$ , Z-score normalized). **c** Pathway differential abundance (DA) analysis in BrMS3 versus BrMS1-2. **d** ActivePathways analysis comparing BrMs to matched primary tumors ( $n = 82$ ) based on integrated RNA-seq and proteomics data. The  $P$ -values were adjusted using Holm's method. **e** Volcano plot of differentially abundant metabolites in lung cancer BrMs versus matched primary tumors ( $n = 12$ ), paired Wilcoxon signed-rank test. **f** Pathway DA analysis in lung cancer BrMs versus matched primary tumors. **g** Schematic of integrative strategy combining VIPER analysis of PDO drug perturbation profiles and primary-BrM proteomics to identify master regulators (MRs), followed by PDO-based validation. Created in BioRender. Wang, X. (2026) <https://BioRender.com/k30j1qd>. **h–k** aREA enrichment plots of the top 30 most activated

(red) and inactivated (blue) MRs identified from BrM-PDOs treated with gamitrinib ( $n = 30$ ) in representative BrMS3-BrM (B282T, **h**) and BrMS1-BrM (B337T, **j**). The  $P$ -values were adjusted using Bonferroni method. Dose-response curves showing cell viability of PDOs derived from B282T (**i**) and B337T (**k**) treated with gamitrinib. **l** Kaplan-Meier analysis of breast cancer BrM patients who underwent radiotherapy (data from Cosgrove et al.<sup>29</sup>), stratified by BrMS subtypes ( $n = 11/7/14/10$ ;  $P = 0.042$ , log-rank test). **m, n** Dose-response curves for one representative BrM-PDO from each BrMS subtype treated with everolimus (**m**) and abemaciclib (**n**). Representative BrM-PDOs were selected based on the availability of complete and high-quality dose-response data or  $IC_{50}$  values close to the median  $IC_{50}$  of the respective subtype. Source data is provided as a Source Data file. VIP variable importance in projection.

relevance of our findings. Higher CTL scores together with elevated expression levels of exhaustion markers suggested that BrMS2 might respond more favorably to cancer immunotherapy. To validate this hypothesis, we re-analyzed 7-color mIF staining images of 12 lung cancer BrMs that were previously published by our group<sup>7</sup> (Fig. 5i). Indeed, BrMS2 displayed increases in PD-L1<sup>+</sup> cells and PD-1<sup>+</sup> T cells, as well as exhibited the highest proportion of Ki-67<sup>+</sup>PD-1<sup>+</sup> T cells, indicating an abundance of proliferating exhausted T cells (Supplementary Fig. 7o–q).

Survival analyses of two independent public datasets (advanced melanoma and NSCLC)<sup>70,71</sup> further supported that patients with a higher BrMS2 signature score superior benefits from immune checkpoint blockade (ICB) therapies ( $P = 0.0015$  and  $P = 0.00085$ ) (Supplementary Fig. 7r–s). It should be noted that these cohorts were not restricted to patients with BrMs.

To further investigate intra-tumor heterogeneity and the cellular composition of the TME across BrMS subtypes, we performed snRNA-seq of fresh frozen craniotomy resected samples derived from 16 patients with BrM, including four patients representing each subtype. This experiment yielded 135,866 nuclei for downstream analysis (Fig. 5j). The unsupervised clustering identified 13 major cell populations (Fig. 5k), annotated using canonical markers (Supplementary Fig. 8a), including four malignant tumor cells (MTC) clusters (c1–c4, epithelial cell markers), two glial cell clusters (oligodendrocytes-OLIG1 and astrocytes-GFAP), two neuron clusters (excitatory neurons-SLC17A7 and interneurons-GAD2), three immune cell clusters (T/NK cells-CD3D/E, B cells-IGHM, and myeloid cells-CIQA/B), and two stromal cell clusters (fibroblasts-COL1A1/2 and endothelial cells-FLT1).

To characterize the TIME landscape of pan-cancer BrMs, we further defined eight T/NK cell subpopulations with classic marker genes (Fig. 5l, Supplementary Fig. 8b). They are NK cells, CD8<sup>+</sup> effector T cell (CD8Teff), CD8<sup>+</sup> exhausted T cell (CD8Tex), CD8<sup>+</sup> progenitor-exhausted T cell (CD8Tpex), CD4<sup>+</sup> naive T cell (CD4Tn), regulatory T cells (Treg), IFN response T cells (Tisg), and proliferating T cells (Tpro). These subsets showed distinct expression patterns of *PDCDI* (*PD-1*), *HAVCR2*, and *TOX* (Fig. 5m–o). The complete list of marker genes is provided in Supplementary Data 21.

Cell composition within each subtype revealed specific TME characteristics (Fig. 5p, q and Supplementary Fig. 8c). BrMS1 contained abundant neural cells and T/NK cells. BrMS2 is mainly characterized by enrichment of T cells and B cells, with most T cells mapped to exhausted CD8 T cells (CD8Tex) and most malignant tumor cells mapped to MTC-c3. Fibroblasts and endothelial cells predominated in BrMS3. MTC-c4 cells were enriched in BrMS4.

The functional enrichment of MTC clusters underscored distinct metabolic and signaling programs (Supplementary Fig. 8d). MTC-c1-Metabolic showed increased mitochondrial respiratory chain complex I (NADH dehydrogenase), suggesting enhanced OXPHOS<sup>72</sup>. MTC-c2-Ionotropic was enriched for ion transport and channel activity. MTC-c3-GTPase was enriched for GTPase regulator and activator functions, suggestive of aberrant Ras/Rho/Rac signaling pathway. MTC-c4-

Proliferative displayed elevated cell cycle, and DNA synthesis and repair, suggesting heightened proliferative capacity. In MTC-c4, the expression levels of several cell cycle-related genes (*CCND1*, *CDK6*, and *E2F5*) were particularly high, consistent with our findings based on the analysis of bulk RNA-seq data (Supplementary Fig. 8e).

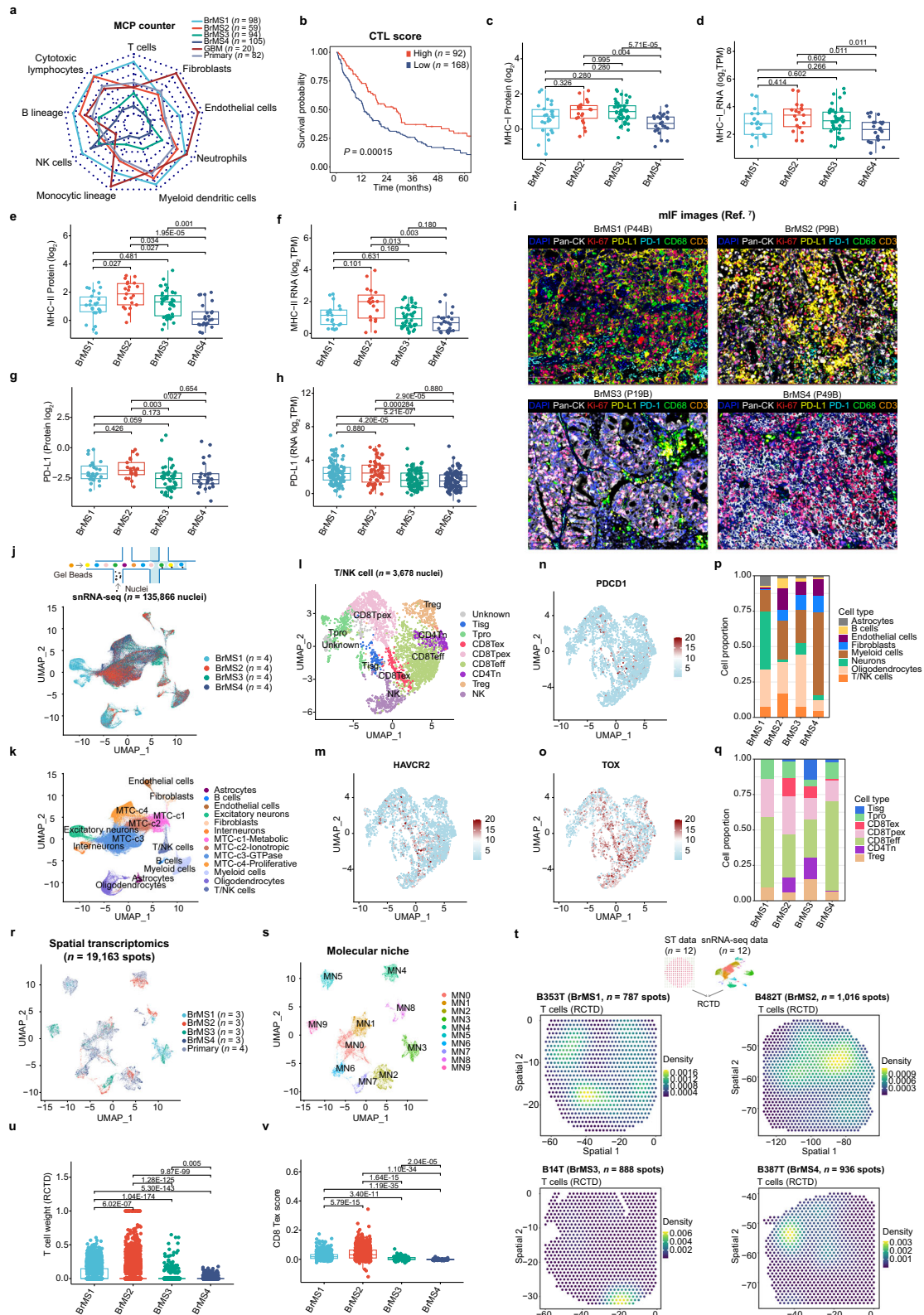
To elucidate intercellular communication network, we applied CellChat analysis<sup>73</sup> across BrMS subtypes (Supplementary Fig. 9a–b). Each subtype showed distinct ligand-receptor signaling. In BrMS1, astrocytes and interneurons had the strongest incoming signals, and *RELN* crosstalk between malignant and neural cells was prominent, supporting a neural-centric phenotype (Supplementary Fig. 9c–e). BrMS2 displayed strong endothelial cells interactions involving *VEGF* and *PECAMI* cross-talks (Supplementary Fig. 9f–g). This finding is well aligned with our previous observation of elevated inflammatory factors (*TNF- $\alpha$* , *IL-6*) in BrMS2, which potentially drive endothelial activation. BrMS2-4 shared enhanced fibroblast interactions, suggestive of active stromal remodeling.

The comparative analysis of macrophage migration inhibitory factor (MIF) signaling revealed striking differences between BrMS3 and BrMS4 (Supplementary Fig. 9h–i). BrMS3 displayed dense crosstalk among malignant cells (c1, c2, c4) and immune cell (myeloid cells, T/NK cells, B cells) or glial cells. In contrast, BrMS4 showed simplified communication limited interactions between MTC-c1 cells and immune cells. Enhanced MIF pathway activity in BrMS3, a crucial pro-inflammatory cytokine, suggested an increased in immune signaling compared to BrMS4<sup>74</sup>. Consistent with its proliferative transcriptomic profile, BrMS4 showed a decrease in immune response, further supporting an inverse relationship between tumor cell proliferation and immune signaling existed<sup>75</sup>.

According to CellChat result, astrocytes were highly activated in BrMS1, while endothelial cells were more activated in BrMS2-4. Thus, we performed NicheNet analysis<sup>76</sup> focusing on these sender-receiver pairs. In BrMS1, astrocyte-derived ligands *TGFB2* and *PTN* showed a strong regulatory potential through their receptors (*TGFBRI/2* and *PTPRB/PTPRZI*) on MTC cells, influencing neuronal and transcriptional plasticity genes (*NRXN3*, *ARHGFI10L*, *ONECUT2*, *FGF12*, *KCNIP4*, *SPOCK1*). These targets are functionally linked to synaptic signaling, EMT/ECM remodeling, and tumor cell plasticity (Supplementary Fig. 10a–c).

BrM2-4 showed endothelial-driven interactions but with distinct downstream responses. In BrMS2, ligands *SEMA3F* and *HSPFG2* regulated ER pathway, ion homeostasis and neuronal signaling related genes (*ESR1*, *NAV2*, *SCNNIA*) (Supplementary Fig. 10d–f). In BrMS3, ligands *SEMA3F* and *ARF1* regulated cytoskeletal and adhesion related genes (*ACTB*, *CDH1*, *SIOOAI10*) (Supplementary Fig. 10g–i). In BrMS4, ligands *VWF* and *EFNB2* regulated ECM remodeling and cell cycle genes (*CCND1*, *COL1A*, *VCAN*) (Supplementary Fig. 10j–l).

To further validate our findings based on the analysis of snRNA-seq data and assess spatial heterogeneity across BrMS subtypes, we conducted spatial transcriptomic profiling of 12 BrM samples (three per subtype) along with four matched primary tumors, identifying 10



distinct molecular niches (MN0-9) (Fig. 5r, s and Supplementary Fig. 11a). Cell type deconvolution by Robust Cell Type Decomposition (RCTD)<sup>77</sup> revealed distinct niche compositions: MN3 was enriched for neuronal cells, MN4 and MN9 for MTCs, MN1 for B cells, MN7 for T cells and myeloid cells, and MN6 for endothelial cells (Supplementary Fig. 11b).

The pathway enrichment analysis further delineated key niches: MN4 and MN9 were enriched in OXPHOS and nucleotide biosynthesis, indicative of a metabolically active phenotype consistent with our results of snRNA-seq data about MTC characteristics (Supplementary Fig. 11c-d). The frequency of niches showed specificity across subtypes: MN3 was predominantly enriched in BrMS1, MN4 was

**Fig. 5 | Immune landscape of BrMs and identification of an immunotherapy-responsive subtype.** **a** Radar plot of immune cell infiltration (MCP counter) across BrMS subtypes, 20 GBMs and matched primary tumors. **b** Kaplan–Meier survival analysis stratified by CTL score (High,  $n = 92$ ; Low,  $n = 168$ ;  $P = 0.00015$ , log-rank test). **c–h** Expression of MHC-I, MHC-II, and PDL1 across BrMS1–4 at protein expression levels ( $n = 26/21/46/25$ ) (**c**, **e**, **g**), at RNA expression levels from newly generated RNA-seq data in the discovery cohort ( $n = 21/19/44/23$ ) (**d**, **f**), at RNA expression levels from the complete discovery cohort ( $n = 98/59/94/105$ ) (**h**). Two-sided Holm-adjusted Wilcoxon rank-sum test, exact  $P$ -values shown. **i** Representative multiplex immunofluorescence images from each BrM subtype (adapted from Duan et al.<sup>7</sup>) ( $n = 12$ , one per subtype shown). Staining includes DAPI (blue), pan-CK (white), Ki-67 (red), PD-L1 (yellow), PD-1 (cyan), CD68 (green), and CD3 (orange). Scale bars: 100  $\mu\text{m}$ . **j** Schematic and UMAP of snRNA-seq data from 16 BrM samples ( $n = 135,866$  nuclei) colored by BrMS subtype. Created in BioRender. Wang, X. (2026) <https://BioRender.com/k3oj1qd>. **k** UMAP of snRNA-seq data from 16 BrM samples colored by cell type annotation. **l** UMAP of T/NK cell subclusters from 16 BrM samples ( $n = 3678$  nuclei). **m–o** UMAP of T/NK cell subclusters from 16

BrM samples ( $n = 3,678$  nuclei) displaying the expression of exhaustion markers HAVCR2 (**m**), PDCD1 (**n**), and TOX (**o**). **p**, **q** Stacked bar plots showing cell type proportions across subtypes from 16 BrM samples. **r** UMAP of ST data from 12 BrM samples and 4 matched primary tumors ( $n = 19,163$  spots) colored by BrMS subtype and primary tumors. **s** UMAP of ST data from 12 BrM samples and 4 matched primary tumors ( $n = 19,163$  spots) annotated by 10 molecular niches (MNO-MN9). **t** Schematic of RCTD and spatial maps showing T cell density in representative samples from each BrMS subtype. Created in BioRender. Wang, X. (2026) <https://BioRender.com/k3oj1qd>. **u**, **v** Boxplots across BrMS subtypes ( $n = 12$ ) showing RCTD-estimated T cell density (**u**  $n = 2,356/3,120/2,691/2,276$  spots) and exhaustion scores (**v**  $n = 764/777/74/98$  spots). Two-sided Holm-adjusted Wilcoxon rank-sum test, exact  $P$ -values shown. In all boxplots, the central line represents the median, and the upper and lower boundaries of the boxes indicate the third quartile (Q3) and the first quartile (Q1), respectively. The upper and lower whiskers extend to the most extreme data points within 1.5 times the interquartile range from the box edges. Source data is provided as a Source Data file.

specifically enriched in BrMS2, MN7 showed high prevalence in BrMS3, and MN6 was abundant in BrMS4 (Supplementary Fig. 11e).

In terms of spatial patterns of T cell distribution across BrMS subtypes, BrMS2 exhibited the highest T cell density with widely distribution, followed by BrMS1 showing localized regions with high density. BrMS3 and BrMS4 demonstrated the lowest density with peripheral T cell clustering (Fig. 5t). Correspondingly, BrMS2 had significantly higher T cell infiltration and exhaustion scores of CD8 T cells than BrMS3 and BrMS4 (Fig. 5u, v), suggestive of a potential benefit from ICB therapies.

The comparative analysis of four BrM-primary pairs uncovered distinct niche distributions. MN3 was specific to BrMs, whereas MN5 and MN1 were more prevalent in primary tumors (Supplementary Fig. 11f–i). The functional enrichment analysis revealed that MN3 was dominant by synaptic signaling pathway, such as modulation of chemical synaptic transmission, regulation of trans-synaptic signaling, and synapse vesicle cycle. MN5 was characterized by small GTPase mediated signaling and membrane homeostasis. MN1 was characterized by extracellular matrix organization and collagen fibril assembly (Supplementary Fig. 11j–l).

Taken together, our results highlight BrMS2 as the subtype most likely to benefit from ICB therapies and underscore distinctions in spatial architectures between BrMs and their primary tumors, revealing BrM specific neuronal signaling niches.

### EMT correlates with immune infiltration in BrMs

The correlation analysis and GSEA revealed that EMT pathway was enriched in proteins positively associated with CTL-high phenotype in BrMs (Fig. 6a–c, Supplementary Data 22). A significant correlation between EMT pathway scores and CTL score was observed only in BrMs ( $\rho = 0.36$ ,  $P = 2.88E-12$ ) but not in primary tumors or GBMs ( $\rho = 0.13$ ,  $P = 0.17$ ) (Supplementary Fig. 12a).

The analysis of an independent cohort of 54 primary tumors and 41 BrMs with NanoString Digital Spatial Profiling (DSP) available confirmed this finding<sup>78</sup>. The mesenchymal marker Fibronectin correlated with cytotoxic markers (GZMB and GZMA), specifically in pan-Cytokeratin positive (pan-CK<sup>+</sup>) BrM regions but not in primary tumors (Fig. 6d, e, Supplementary Fig. 12b).

Using RNA-seq data from the discovery cohort, five EMT gene signatures (MP12-EMT-I, MP13-EMT-II, MP14-EMT-III, MP15-EMT-IV and MP16-MES)<sup>79</sup> showed strong positive correlations with immune infiltration specifically in BrMs but not in primary tumors (Supplementary Fig. 12c). The association remains significant after we adjusted for fibroblast and endothelial cell fractions (Supplementary Data 23). These were further validated in three independent validation cohorts (Supplementary Fig. 12d). EMT scores were higher in BrMS1 and BrMS2 subtypes than in BrMS3 and BrMS4 from discovery cohort (Supplementary Fig. 12e).

At the single cell level, we identified MTCs and CD8 T cells from 16 BrM samples and revealed a positive correlation between EMT pathway activity from MTCs and cytotoxic lymphocytes scores in CD8 T cells ( $\rho = 0.65$ ,  $P = 0.0079$ ) (Supplementary Fig. 12f). The mIF of six newly collected pan-cancer BrM samples in this study supported this relationship, showing a significant decrease in spatial separation between CD8<sup>+</sup> T cells and pan-CK tumor regions compared with pan-CK<sup>+</sup> regions (Supplementary Fig. 12g–h).

With spatial transcriptomic profiling, regions were classified as T-neighbored Tumor if they contained enriched T cell populations or were adjacent to T cell enriched spots, whereas regions lacking T cell enrichment both internally and in adjacent spots were defined as non-T-neighbored Tumor (Fig. 6f, g). The differential expression analysis revealed a significant enrichment of EMT pathway in T-neighbored Tumor regions (Fig. 6h–j). Key EMT associated genes, including *IL32*, *CXCL12*, *VCAM1*, and *NNMT*, showed higher expression in T-neighbored Tumor regions (Fig. 6k). These findings suggested a spatial relationship between EMT activity and T cell infiltration.

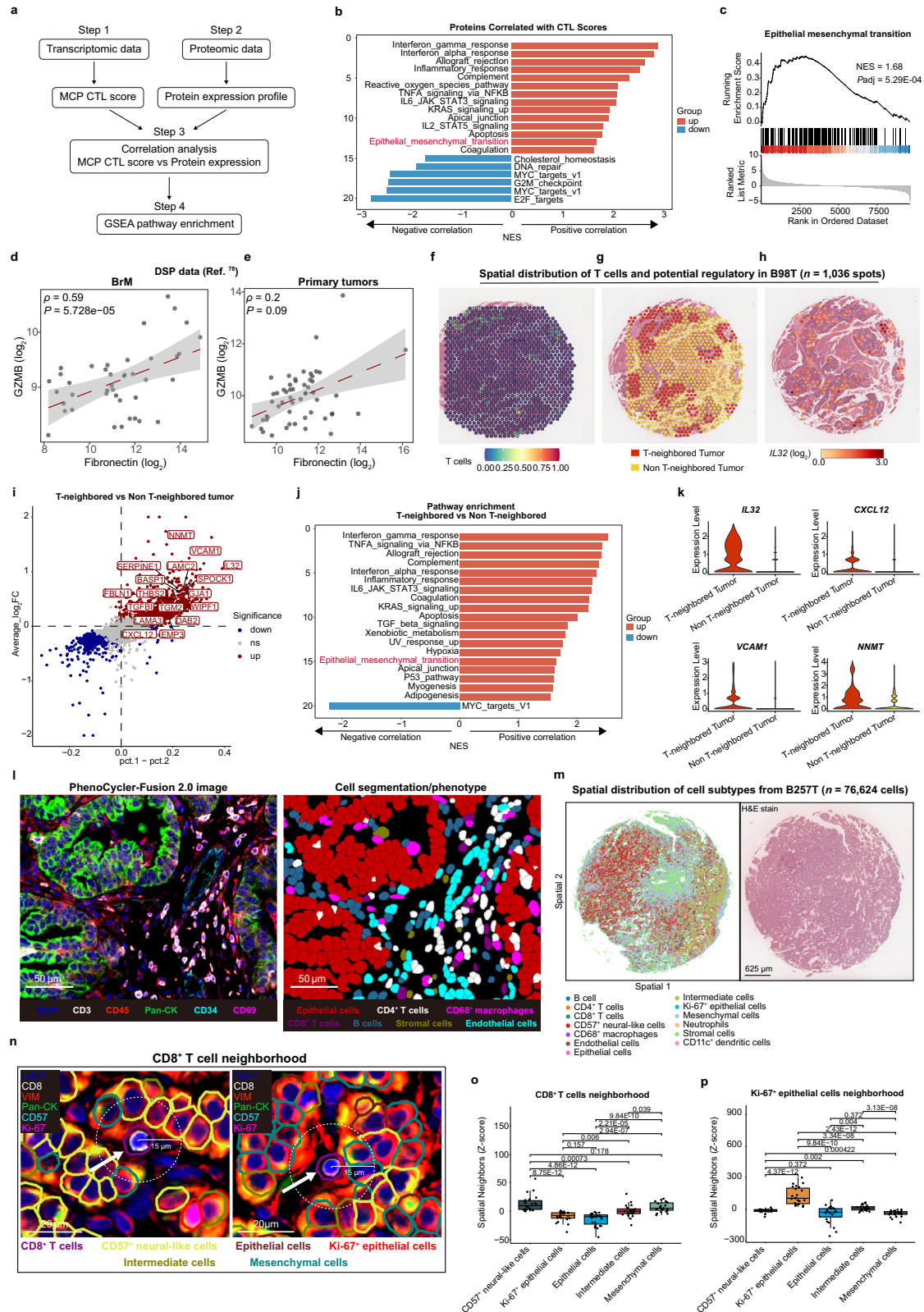
To validate these spatial architectures and specifically visualize the proximity between EMT-high tumor cells and T cells at single-cell resolution, we performed PhenoCycler immuno-oncology (IO) 60 immune panel on 23 BrMs (Fig. 6l, Supplementary Fig. 12i). Ten major cell types were identified, including endothelial cells, B cells, neutrophils, CD68<sup>+</sup> macrophages, CD11c<sup>+</sup> dendritic cells, CD8<sup>+</sup> T cells, CD4<sup>+</sup> T cells, malignant cells, NK cells, and stromal cells (Supplementary Fig. 12j). Malignant cells were further classified into epithelial cells, intermediate cells, Ki-67<sup>+</sup> epithelial cells, mesenchymal cells, and CD57<sup>+</sup> neural-like cells (Supplementary Fig. 12k). The spatial distribution of these cell types was visualized in a representative sample (B257T) (Fig. 6m).

Neighborhood analysis within a 15  $\mu\text{m}$  radius on CD8<sup>+</sup> T cells and CD4<sup>+</sup> T cells revealed mesenchymal and CD57<sup>+</sup> neural-like cells were the most abundant cell types (Fig. 6n, o, Supplementary Fig. 12l–m). CD57<sup>+</sup> neural-like cells, which indicate neural tumors<sup>80</sup>, along with mesenchymal tumor cells, may promote immune infiltration. Additionally, Ki-67<sup>+</sup> epithelial cells demonstrated strong self-clustering, indicating high localized proliferation characteristics (Fig. 6p, Supplementary Fig. 12n).

Taken together, our consistent results from bulk, single-cell and spatial level suggest that EMT is actively associated with immune infiltration in BrMs, potentially facilitating the T cell engagement within the brain metastatic microenvironment.

## Discussion

In this study, we have assembled and analyzed a multi-omic atlas of 1032 pan-cancer BrMs, integrating with genomic, transcriptomic, proteomic, metabolomic, single-nucleus, and spatial profiling data.



This comprehensive resource reveals four robust molecular subtypes, including BrMS1 (neural-like), BrMS2 (immune-infiltrated with coordinated EMT), BrMS3 (metabolic), and BrMS4 (proliferative and genomically unstable) that transcend primary tumor origins and correlate with distinct clinical outcomes, cellular architectures, and therapeutic vulnerabilities. By validating these subtypes across independent cohorts and platforms, and confirming their biological relevance

through PDOs and functional assays, we establish a classification framework that shifts the paradigm from primary tumor-driven categorization to one centered on metastatic adaptation in the brain<sup>1,2</sup>.

The utilization of 4D-DIA proteomics provides a comprehensive proteomic landscape that complements the transcriptomic-based classification, shedding light on the molecular underpinnings of the identified subtypes. The subtype-specific protein signatures captured

**Fig. 6 | EMT Correlates with immune infiltration in BrMs.** **a** Workflow for correlating transcriptomic MCP CTL score with protein expression in 118 cases with paired RNA-seq and proteomic data. **b** GSEA of proteins positively and negatively correlated with the MCP CTL score. **c** GSEA enrichment plot for EMT signature ( $n = 118$ ). NES and BH-adjusted  $P$ -values are indicated. **d, e** Spearman correlation (two-sided) between EMT marker Fibronectin (FNI) and cytotoxic T cell activity marker GZMB expression in BrM ( $n = 41$ ) and primary tumors ( $n = 54$ ) using a DSP dataset from Schoenfeld et al.<sup>78</sup>. The gray shaded area in the plot represents the 95% CI for the regression line. **f-h** Spatial analysis of ST data from one sample B98T ( $n = 1,036$  spots): T cell density estimated by RCTD (**f**), classification of tumor spots into T-neighborhood (red) and non-T-neighborhood (yellow) (**g**), spatial quantification of *IL32* (**h**). **i-k** Differential gene expression analysis between T-neighborhood and Non-T-neighborhood Tumor regions using ST data from 12 BrM samples: Volcano plot showing differentially expressed genes (**i**), GSEA of hallmark pathways showing top

20 enriched pathways in T-neighborhood regions with EMT pathway highlighted (**j**), violin plots showing expression levels of key EMT-associated genes (*IL32*, *CXCL12*, *VCAM1*, and *NNMT*) (**k**). **l** Representative PhenoCycler-Fusion image showing spatial distribution of immune, stromal, endothelial, and malignant cells, with corresponding raw imaging (left) and phenotyping (right). **m** Spatial distribution of cell types and H&E imaging in one representative sample (B257T,  $n = 76,624$  cells). **n** Spatial neighborhood between CD8<sup>+</sup> T cells and tumor cell subtypes within a 15  $\mu$ m (white dashed circles). **o, p** Quantification of spatial neighbors within CD8<sup>+</sup> T cell neighborhoods (**o**) and Ki-67<sup>+</sup> epithelial cells neighborhoods (**p**) ( $n = 23$ ; two-sided t-test, exact  $P$ -values shown). The central line represents the median, and the upper and lower boundaries of the boxes indicate the third quartile (Q3) and the first quartile (Q1), respectively. The upper and lower whiskers extend to the most extreme data points within 1.5 times the interquartile range from the box edges. Source data is provided as a Source Data file.

here, ranging from glial and synaptic factors in BrMS1 to the prominent immune- and inflammation-related proteins in BrMS2, epithelial and metabolic markers in BrMS3, and proliferative and cell cycle regulators in BrMS4, emphasize the multifaceted nature of BrM biology<sup>81–83</sup>. These findings are consistent with our transcriptomic observations and highlight the importance of integrating multiple omics layers for a comprehensive molecular characterization<sup>3,84,85</sup>. Moreover, the observed mRNA-protein correlation patterns, which differed significantly among subtypes, underscore the complexity of regulatory mechanisms in BrMs and the necessity of incorporating proteomic data to move beyond transcript-level inference.

Pathway analysis using both transcriptomic and proteomic data revealed subtype-specific biological processes, providing critical insights into their distinct pathophysiology. The enrichment of neuronal signaling pathways in BrMS1 and the immune and stromal remodeling in BrMS2 aligns with previous evidence showing that BrMs co-opt local brain cells and leverage inflammatory networks to facilitate tumor growth<sup>86</sup>. Similarly, the metabolic reprogramming in BrMS3 and the heightened proliferative signaling in BrMS4 are in line with established oncogenic paradigms that adapt primary tumor traits to the unique brain microenvironment<sup>87,88</sup>. The WGCNA based identification of prognostically significant co-expression modules, such as ME21 and ME30, pinpoints EMT related and OXPHOS function related protein clusters as potential molecular targets with direct clinical relevance. The negative impact on survival associated with heightened proliferative signaling highlights the importance of these pathways as deleterious in the setting of BrMs<sup>3,86</sup>.

The comprehensive genomic profiling of pan-cancer BrMs elucidates distinct mutational landscapes across the four identified molecular subtypes, highlighting the underlying molecular heterogeneity and potential therapeutic vulnerabilities inherent in BrMs. Notably, the BrMS1 is characterized by mutations in key immune regulatory genes such as *KDM2A* and *TRAF2*, which aligns with its enhanced immune cell infiltration and lower genomic instability, consistent with the findings of Brastianos et al.<sup>3</sup>, which emphasized the role of the TiME in BrM progression. The prevalence of *TP53* mutations across all subtypes, particularly in BrMS4, underscores its pivotal role as a driver in BrMs, echoing its established function in various cancers<sup>89</sup>. The Signature 1, indicative of age-related mutational processes, and Signature 4, associated with APOBEC-mediated mutagenesis, reflect distinct mutational mechanisms that may influence tumor behavior and therapeutic responses<sup>90</sup>.

The integration of snRNA-seq and spatial transcriptomic data provides valuable insights into the complex cellular ecosystems that shape the TME of BrMs and further refines the characterization of the identified molecular subtypes. Our observation that malignant and non-malignant cells from distinct subtypes cluster together in a transcriptome-based UMAP space suggests that certain transcriptional programs may transcend subtype boundaries, while still allowing for

distinct subtype-specific cellular niches and signaling patterns. These findings are consistent with previous work demonstrating that spatially and functionally specialized cells can converge on similar molecular states, reflecting convergent adaptations to the unique conditions of the brain metastatic environment<sup>85</sup>. The identification of diverse cell populations and molecular niches underscores the remarkable complexity of BrMs and highlights the crucial interplay between tumor cells and their surrounding milieu<sup>86</sup>.

Our integrated analyses demonstrate that the TME in the brain profoundly shapes BrM biology, driving convergent phenotypes across diverse primary origins<sup>85–88,91</sup>. Unlike primary tumors or extracranial metastases, BrMs exhibit heightened chromosomal instability<sup>92–94</sup>, metabolic reprogramming (e.g., enhanced OXPHOS and TCA cycle activity)<sup>5,7,8</sup>, and subtype-specific immune landscapes<sup>7,86,95</sup>. BrMS1 and BrMS2 represent “immune-hot” states, characterized by elevated CTL infiltration and MHC expression, while BrMS3 and BrMS4 are “immune-cold,” with suppressed antigen presentation and lower CTL scores. ST and mIF reveal how these states manifest architecturally: BrMS2 tumors display diffuse T cell infiltration with high PD-1/PD-L1 expression and proliferating exhausted T cells, suggesting an inflamed yet checkpoint-suppressed niche<sup>7,78</sup>. These findings align with emerging evidence that the brain’s unique milieu, rich in astrocytes, neurons, and limited vascular permeability, selects for adaptive strategies: immune evasion through proliferation in BrMS4 versus co-opting inflammatory signaling in BrMS2. This TME-driven convergence explains why primary tumor subtypes poorly predict BrM behavior and underscores the need for metastasis-specific classifications<sup>1,6</sup>.

A key discovery is the positive correlation between EMT and immune infiltration in BrMs, contrasting with the immunosuppressive role of EMT in many primary tumors. Proteomic and transcriptomic analyses show that EMT-related proteins and gene signatures (e.g., MPI2-EMT-I, MPI6-MES) strongly associate with CTL abundance and cytolytic activity specifically in BrMs. Spatial profiling via mIF and NanoString DSP confirms this: mesenchymal markers like Fibronectin co-localize with cytotoxic granzymes in pan-CK<sup>+</sup> regions, while epithelial zones (pan-CK<sup>-</sup>) show greater distances to CD8<sup>+</sup> T cells. At the single-cell level, EMT meta-programs cluster with immune response programs (e.g., MHC, interferon), suggesting coordinated regulation. We propose a testable model: in the CNS, EMT-driven extracellular matrix remodeling creates a “permissive scaffold” for T cell trafficking, potentially by degrading dense glial barriers or recruiting chemokines (e.g., CXCL8 in BrMS2). It is important to emphasize that, unlike GBM which exhibit extensive parenchymal invasion driven by CNS-intrinsic glial properties, the EMT program in BrMs primarily mediates local microenvironmental remodeling and immune niche formation rather than enhancing macroscopic invasive capacity. This EMT-immune interaction is absent in matched primary tumors and GBMs as well. Intriguingly, this immune-permissive phenotype could explain why

EMT-high BrMs (BrMS1/2) respond better to radiotherapy or ICB. Future studies using CRISPR-edited PDOs or intravital imaging could validate this hypothesis and identify EMT modulators as immune-sensitizing agents.

PDO experiments nominate subtype-linked molecular targets and pathway dependencies that are consistent with the multi-omic features defined in BrMS subtypes. BrMS3's metabolic rewiring, marked by upregulated glycolytic and nucleotide intermediates, implicates dependence on the mTOR axis, consistent with enriched fatty acid metabolism and ER stress pathways. BrMS4's proliferative hallmarks, including cell cycle activation and genomic instability (e.g., high SCNA, TP53/RB1 alterations), indicate vulnerability of the CDK4/6 axis, aligning with elevated E2F/FOXM1 networks. Given the limited statistical power within each subtype, the current findings should be regarded as exploratory and hypothesis-generating. For immune-hot subtypes, BrMS1 shows superior radiotherapy response, likely due to neural niche radiosensitivity<sup>29,60</sup>, while BrMS2's exhausted T-cell profile predicts ICB benefit, as evidenced by improved survival in melanoma and NSCLC cohorts<sup>16–23,70,71</sup>. However, important limitations must be acknowledged. These cohorts were not restricted to patients with BrMs. The response to immunotherapy may differ significantly between extracranial sites and BrM due to the unique immune microenvironment of the central nervous system and the presence of the BBB. Therefore, further validation studies specifically focusing on BrM patients with systemic brain imaging and BrM-specific endpoints are needed to confirm whether the BrMS2's favorable response to ICB can be extended to the brain metastatic setting. To maximize impact, we advocate rational combinations: for BrMS4, pairing CDK4/6 inhibitors with immunomodulators (e.g., STING agonists) to disrupt compact clustering and convert “cold” TMEs to “hot”<sup>96</sup>; for BrMS3, combining mTOR inhibitors with mitochondrial-targeted agents like gamitrinib to exploit OXPHOS dependency<sup>7</sup>. These strategies, grounded in our multi-omic roadmap, could guide patient stratification in clinical trials, potentially improving outcomes in this heterogeneous disease<sup>1,2</sup>.

While our study represents an advance to understand BrM biology, several limitations warrant consideration. Treatment covariates were not included as these data were unavailable for most patients. Our molecular classification reflects the tumor's biological state at resection, though we cannot exclude that prognostic differences may be partially influenced by differential treatment responses. One limitation of our drug sensitivity assays is that technical replicates were not included, which could affect the accuracy and reproducibility of the measured responses. Incorporating replicate wells in future assays will strengthen the reliability of these findings. Sample representation biases towards common primaries (e.g., lung, breast) may underpower rarer origins, and metabolomic coverage, though targeted, could miss niche pathways.

While our study reveals potential actionable targets for distinct molecular subtypes, it is important to emphasize that its core contribution is to elucidate the underlying biological programs, rather than to provide clinical management recommendations. Therefore, future research should prioritize validating these subtype-target associations in prospective cohorts and functionally exploring them in more complex preclinical models, which are critical steps before any consideration for clinical translation.

In conclusion, this work establishes a paradigm for understanding brain metastasis, shifting the focus from a primary tumor-centric view to one defined by the biological states adopted within the brain itself. This spatially resolved and functionally validated atlas provides not only a foundational resource but also a biological and conceptual roadmap for future investigation. By linking molecular subtypes to specific immune architectures and actionable vulnerabilities, we provide a foundation for stratifying future research and for developing subtype-informed therapeutic hypotheses to address this devastating disease.

## Methods

### Patient cohorts

This multicenter study incorporated both discovery and validation cohorts. The discovery cohort comprised a newly generated cohort from QMH and SYSUCC, complemented by data from five previously published cohorts<sup>5,7,8,25,26</sup>. For external validation, we assembled an independent cohort consisting of newly enrolled patients from BJTTH, SYSUCC, and WCH, which is referred to throughout this study as the In-house Validation cohort, along with data derived from ten published cohorts. We systematically retrieved medical records and archived tissue specimens. Sex was determined based on biological attributes in medical records. Gender identity information was not specifically collected. Due to ethical restrictions, individual sex information was not available, and age was provided as four ranges (< 50, 50–59, 60–69, and ≥ 70 years). Written informed consent was obtained from all participants for their samples and clinical information to be used in this study. The collection of samples at QMH was approved by the Institutional Review Board of the University of Hong Kong Hospital Authority Hong Kong West Cluster (HKU/HA HKWIRB, IRB Reference Number: UW07-273). Sample collection was approved by its Institutional Review Board/Independent Ethics Committee at SYSUCC (Protocol B2021-256-Y02), BJTTH (KY2024-100-01) and WCH (2023-1762). All procedures adhered to the Declaration of Helsinki and Good Clinical Practice Guidelines.

### Sample processing

Tissues were harvested and immediately fresh frozen in liquid nitrogen. Tissue samples were homogenized using the TissueLyser II (Qiagen) in 350 µl Buffer RLT Plus with 5 mm stainless steel beads. Then genomic DNA and total RNA were simultaneously purified from the same lysate using the AllPrep DNA/RNA Micro Kit (Qiagen) according to the manufacturer's protocol. Briefly, the lysate was first passed through an AllPrep DNA spin column to selectively bind DNA. Ethanol was added to the flow-through and the sample was applied to a RNeasy MinElute spin column to bind total RNA. Both columns were washed, and the nucleic acids were eluted. For very small samples (< 2 µg tissue), 5 µg poly-A carrier RNA (provided in the kit) was added to the lysate. The quantity and quality of the purified DNA and RNA were assessed by UV spectrophotometry and agarose gel electrophoresis. The purified genomic DNA and total RNA were ready for immediate use in downstream applications or stored at –20 °C and –80 °C, respectively, for later analysis.

### Bulk RNA sequencing and subtype classification

#### Library construction for RNA-seq and sequencing procedures.

Total RNA was isolated using the AllPrep DNA/RNA Micro Kit (Qiagen). Paired-end libraries were synthesized using the TruSeq<sup>®</sup> RNA Sample Preparation Kit (Illumina, USA), following the TruSeq<sup>®</sup> RNA Sample Preparation Guide. Poly-A-containing molecules were purified with poly-T oligo-attached magnetic beads. The purified mRNA was fragmented into small pieces using divalent cations at 94 °C for 8 min. The cleaved RNA fragments were converted into first-strand cDNA using reverse transcriptase and random primers, followed by second-strand cDNA synthesis using DNA Polymerase I and RNase H. These cDNA fragments underwent end repair, addition of a single 'A' base, and adapter ligation. The products were then purified and enriched by PCR to create the final cDNA library. The purified libraries were quantified using a Qubit<sup>®</sup> 2.0 Fluorometer (Life Technologies, USA) and validated with an Agilent 2100 bioanalyzer (Agilent Technologies, USA) to confirm the insert size and calculate the molar concentration. Clusters were generated on a cBot with the library diluted to 10 pM, followed by sequencing on the Illumina HiSeq Xten (Illumina, USA).

**The analysis of bulk tumor RNA-seq data.** Raw sequencing reads were subjected to QC and adapter trimming using Trimmomatic

(v0.39). The processed high-quality reads were subsequently aligned to the human reference genome (hg38) using HISAT2 (v2.1.0)<sup>97</sup>. Following alignment, StringTie (v1.3.4 d) was employed to quantify gene expression levels, generating both transcripts per million (TPM) and raw count matrices based on the reference annotation.

To ensure data quality, genes were retained only if they were expressed in more than 50% of samples across each cohort. Batch effects were addressed using the Combat-seq for count data, implementing a negative binomial regression model<sup>39</sup>. For TPM data and processed microarray data, the Combat algorithm was applied. Missing values were imputed by KNN method. Principal component analysis (PCA) was performed to evaluate the effectiveness of batch effect removal. Differential expression analysis was conducted using DESeq2, which fits a negative binomial model for each gene<sup>97</sup>. Genes with  $FDR \leq 0.05$  and absolute  $\log_2$  fold change ( $|\log_2FC|$ )  $> 1$  were considered significantly differentially expressed.

In a nCounter data with a panel of 770 pan-cancer immune genes (External Validation 3), the expression values of BrMS1-3 signatures were calculated by Z-scores and min-max normalized. BrMS4 was not calculated because only two genes were available in this panel. Samples were classified as BrMS4 if BrMS1-3 scores lower than the threshold estimated by Contal and O'Quigley method<sup>98</sup>.

**Unsupervised clustering of transcriptomic data.** To identify molecular subtypes across pan-cancer BrMs, we selected the top 4000 most variable genes based on standard deviation for unsupervised clustering analysis. Consensus clustering was implemented using the ConsensusClusterPlus R package (v1.66.0)<sup>40</sup> with the following parameters: 1000 bootstrap repetitions, 80% sample resampling rate ( $pltem = 0.8$ ), Euclidean distance metric, and k-means clustering algorithm evaluating up to 10 clusters. The optimal number of clusters was determined to be four ( $k = 4$ ) based on an analysis of cluster stability and prognostic power across different k values, which maximized both the Wald test statistic for survival differences and the C-index for prognostic discrimination. Prediction strength (fpc v2.2-12) were also used for the selection of best k value. We also run the consensus clustering in lung cancer, melanoma and breast cancer subgroups to confirm the selection of  $k = 4$  is optimistic.

**Validation of molecular subtypes.** To characterize subtype-specific molecular signatures, we first identified core samples within each subtype by selecting those with positive silhouette width scores. Differential expression analysis was performed in core samples using DESeq2 to compare each subtype against the remaining three subtypes. Signature genes for each subtype were defined as uniquely upregulated genes ( $\log_2FC > 1$ , adjusted  $P < 0.05$ ), resulting in 711, 240, 262, and 264 signature genes for BrMS1, BrMS2, BrMS3, and BrMS4, respectively. The robustness of the four-subtype classification was validated using an integrated external dataset comprising 521 samples from published pan-cancer BrMs studies (RNA-seq and microarray) and a newly generated RNA-seq cohort from three hospitals. NTP was performed using the subtype-specific signature genes with 1000 permutations after data normalization and scaling<sup>41</sup>. Samples with  $FDR < 0.05$  were classified as unknown subtype. To build a classifier for the four molecular subtypes, we employed a machine learning approach using an elastic-net penalized logistic regression model. We first calculated single-sample pathway scores based on Reactome gene sets using the ssMWW-GST method. The model was built using the R package glmnet (v4.1-10) and trained using pathway enrichment features, including NES and FDR values. Hyperparameters (alpha and lambda) were tuned via 10-fold cross-validation to optimize model performance. Samples were randomly divided into a training cohort (70%) and a validation cohort (30%). Model performance was evaluated using accuracy and the ARI.

**Kaplan–Meier survival analysis.** Post-operative survival was defined as the time interval between the craniotomy for BrM and cancer-related death or the date of the last follow-up, whichever occurred first. Patients with available survival data were included in Kaplan–Meier analyses. Survival comparisons among clusters were performed using Cox proportional hazard models and log-rank tests implemented through the survminer package (v0.4.9). Multivariable Cox proportional hazard models were built to estimate the association between several clinical factors and patients' survival. Kaplan–Meier analysis was employed for survival curve generation, and forest plots were constructed using the R forestplot package for comprehensive visualization of survival outcomes.

**Functional enrichment analysis.** Comprehensive functional annotation of the identified clusters was performed using the clusterProfiler package (v4.4.4)<sup>99</sup>. We employed both hypergeometric testing (enrichr function) and gene set enrichment analysis (GSEA function). Multiple curated gene sets from MSigDb were utilized, including Gene Ontology (GO) terms, Hallmark gene sets, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways, and Reactome pathways, to provide a thorough characterization of the biological processes associated with each molecular subtype<sup>43</sup>.

**IHC staining.** Tissue sections from formalin-fixed, paraffin-embedded specimens underwent systematic deparaffinization: initial heating at 65 °C for 30 min, followed by sequential immersion in biological clearing agent (2 × 10 min), absolute ethanol (5 min), and descending ethanol gradients (95%, 90%, and 80%; 2 min each), concluding with a 2-min double-distilled water rinse.

Antigen retrieval was performed using either citrate buffer (pH 6.0) or EDTA buffer (pH 8.0/9.0) with microwave treatment (two 8-min cycles). Endogenous peroxidase activity was quenched using 3% H<sub>2</sub>O<sub>2</sub> in methanol (30 min, room temperature). Following washing steps, sections were blocked with goat serum (30 min, 37 °C) and subsequently incubated with anti-GFAP (CST, 3670, 1:200), anti-IF116 (Abcam, ab169788, 1:200), anti-ACSL5 (HUABIO, HA601166, 1:1000), and anti-TOP2A (Abcam, ab52934, 1:1000) antibodies overnight at 4 °C. HRP-conjugated secondary antibody incubation was conducted at 37 °C for 1 h. Immunoreactivity was detected using DAB chromogen and counterstained with hematoxylin. The stained sections underwent dehydration and clearing before mounting with neutral rapid-drying medium for microscopic evaluation.

**Transcriptional factor analysis.** Transcription factor activity analysis was performed using Cancer Core Transcription Factor Specificity (CaCTS, v1.0) to compare regulatory patterns across the four subtypes<sup>100</sup>. Representative samples for each tumor subtype were generated by calculating mean expression values for individual transcription factors. The CaCTS score was utilized to quantify gene specificity within each subtype. Additionally, regulon activity was assessed using the SCENIC (Single-cell Regulatory Network Inference and Clustering) workflow<sup>101</sup>. This process involved initial inference of co-expression modules between transcription factors and candidate target genes using GENIE3, followed by cis-regulatory motif analysis via RcisTarget to identify modules with significant transcription factor motif enrichment. Regulon activity scoring was performed using the AUCell algorithm, and master regulon specificity was quantified using Jensen-Shannon Divergence (JSD).

**Stemness score inference.** We used stemness score to assess the degree of oncogenic dedifferentiation as previously described<sup>102</sup>. The analysis utilized pluripotent stem cell samples (ESC and iPSC) from the Progenitor Cell Biology Consortium (PCBC) dataset (Synapse ID syn4976369)<sup>103</sup>. A predictive model was constructed using the one-class logistic regression (OCLR) machine-learning algorithm on the

PCBC dataset, which was subsequently applied to our dataset for stemness score prediction.

#### 4D-DIA quantitative proteomics and co-expression network analysis

**Protein extraction, digestion, and cleanup.** 107 BrMs with 3 matched primary tumors and 20 GBMs were performed by 4D-DIA proteomics. First, fresh frozen tissue was centrifuged at 15,000 g at 4 °C for 10 min, then lysed with a cracking buffer containing 1 mM PMSF and 2 mM EDTA for 5 min, followed by instantaneous centrifugation; the resulting protein was quantified using a BCA quantitative kit, and 10 µg of this protein solution was processed using the SISPROT kit (Shenzhen Beipuo Biotechnology Co., Ltd.) to obtain peptides, which were then quantified with a BCA kit and analyzed by mass spectrometry after lyophilization. Equal quantities of proteins from each sample underwent tryptic digestion beginning with the addition of 8 M urea in 200 µl to the supernatants. The mixture was then reduced with 10 mM DTT at 37 °C for 45 min and alkylated with 50 mM iodoacetamide (IAM) for 15 min in a dark room at room temperature. Subsequently, four volumes of chilled acetone were added, and the samples were precipitated at -20 °C for 2 h. After centrifugation, the protein precipitate was air-dried and resuspended in 200 µl of 25 mM ammonium bicarbonate solution, with 3 µl of trypsin (Promega) added and digested overnight at 37 °C. Post-digestion, the peptides were desalted using a C18 Cartridge, dried using a vacuum concentrator, concentrated by vacuum centrifugation, and finally redissolved in 0.1% (v/v) formic acid for analysis.

**LC-MS/MS analysis.** Liquid chromatography (LC) was conducted using a nanoElute UHPLC system (Bruker Daltonics, Germany). Approximately 200 ng of peptides were separated over 60 min at a flow rate of 0.3 µl/min using a commercially available reverse-phase C18 column equipped with an integrated CaptiveSpray Emitter (25 cm × 75 µm ID, 1.6 µm, Aurora Series with CSI, IonOpticks, Australia). The separation temperature was maintained at 50 °C by an integrated Toaster column oven. Mobile phases A and B consisted of 0.1 vol.-% formic acid in water and 0.1% formic acid in acetonitrile (ACN), respectively. The gradient of mobile phase B was adjusted from 2% to 22% over the first 45 min, increased to 35% over the subsequent 5 min, further increased to 80% over the next 5 min, and then held at 80% for an additional 5 min. The LC system was coupled online to a hybrid timsTOF Pro2 (Bruker Daltonics, Germany) through a CaptiveSpray nano-electrospray ion source (CSI). The timsTOF Pro2 operated in Data-Dependent Parallel Accumulation-Serial Fragmentation (PASEF) mode, capturing 10 PASEF MS/MS frames within 1 complete frame. The capillary voltage was set at 1400 V, acquiring MS and MS/MS spectra from 100 to 1700 m/z. The ion mobility range (1/KO) was set between 0.7 to 1.4 Vs/cm<sup>2</sup>. Both the TIMS accumulation and ramp time were maintained at 100 ms, facilitating operation at duty cycles close to 100%. A “target value” of 10,000 was employed with a repeated schedule, and the intensity threshold was established at 2500. Collision energy was linearly ramped based on mobility, from 59 eV at 1/KO = 1.6 Vs/cm<sup>2</sup> to 20 eV at 1/KO = 0.6 Vs/cm<sup>2</sup>. The quadrupole isolation width was set to 2Th for m/z < 700 and 3Th for m/z > 800, optimizing the system for precise mass isolation and fragmentation.

**Database search and quantification.** Raw data were processed using DIA-NN (v1.8.1) in library-free mode against the UniProt human reference proteome database (UP000005640, downloaded on 2023-05-04, 82,492 entries) supplemented with iRT2 standard sequences. Spectral libraries were generated using deep learning-based prediction algorithms with default DIA-NN parameters. Match-between-runs (MBR) was enabled to improve quantification coverage. Protein quantification was performed using the MaxLFQ algorithm. Results were filtered at 1% FDR at both precursor and protein levels.

**The analysis of 4D-DIA data.** Multiple batch integration was performed after filtering proteins with > 50% missing values. Data underwent log-normalization and batch effect correction using ComBat. Missing values were imputed using DreamAI ensemble algorithm. Differential protein expression analysis employed Wilcoxon rank-sum tests (significance criteria: |logFC| > 0.58, *P* < 0.01).

**WGCNA.** WGCNA was performed on 9621 proteins across 138 tumor samples using WGCNA R package (v1.72-1) (power = 14, *R*<sup>2</sup> = 0.85, minimum module size = 50). Module functional annotation was performed using clusterProfiler with MSigDB-derived gene sets, including Hallmark, GO-BP, and KEGG. Representative pathways were selected based on adjusted *P*-values combined with gene counts. Module eigengene values were calculated as the first principal component of gene expression patterns within each module. Network visualization was accomplished using igraph package. Networks were analyzed in Cytoscape (v3.10.4) using the cytoHubba plugin. The top ten proteins were ranked by Maximal Clique Centrality (MCC).

**Integrative analysis of transcriptomics and proteomics data.** Partial Spearman correlation analysis, adjusting the tumor purity, between transcriptomic and proteomic data utilized 8450 genes with > 50% completeness across both modalities. Pathway enrichment analysis was performed with ActivePathways (v2.0.5) using a gene × omics *P*-value matrix (RNA-seq differential expression and proteomic differential abundance [DA]) and visualized via aPEAR package<sup>46,104</sup>. At the gene level, we applied directional *P*-value merging (DPM) with a pre-specified concordant direction *c* (1,1). Direction was defined by effect-size cutoffs: RNA logFC > 1 and *P* < 0.05, protein logFC > 0.58 and *P* < 0.05. All *P*-values were merged and adjusted by DPM methods as mentioned in ref. 46. Integrative pathway enrichment analysis was implemented in the ActivePathways R package. Holm’s method was used for multiple testing correction at the pathway level. Multi-omics factor analysis (MOFA2, v1.16.0)<sup>57</sup> was performed as described based on a subset of the series with matched RNA-seq, proteomic and genomic measurements. The number of input features for each modality was 13,336 (RNA), 9,621 (protein), and 59 (CNV), respectively. 89 samples profiled with all three data type were used for modal training. The number of latent factors was set to num\_factors = 10, convergence\_mode = ‘slow’, while all other parameters were kept at their default values. The associations between 10 factors and external covariates were calculated by correlate\_factors\_with\_covariates function. GSEA of genes or proteins ranked by latent factor scores was performed as previously described<sup>105</sup>.

#### Whole-exome sequencing and integrative analysis of multi-omics data

**Library construction for whole exome sequencing and sequencing procedures.** DNA samples were first evaluated for QC. Each sample required 1 µg of DNA (quantified by Qubit), with integrity verified by agarose gel electrophoresis to ensure no degradation or RNA contamination. DNA purity was assessed using Nanodrop spectrophotometry (OD<sub>260/280</sub> = 1.8 - 2.0). Samples with less than 1 µg or showing degradation underwent modified library preparation protocols following client approval.

Library construction was performed using the Agilent SureSelect Human All Exon V6 kit according to the manufacturer’s protocol by Gaojing (Zhejiang Anji) Precision Medicine Technology Co., Ltd. Briefly, qualified genomic DNA was randomly fragmented to 150–300 bp, followed by end repair, 3’ adenylation, and paired-end adapter ligation. The pooled libraries were then hybridized with biotinylated probes for exome capture, and target regions were isolated using streptavidin magnetic beads. The captured libraries underwent PCR amplification and quality assessment.

Sequencing was performed on the Illumina NovaSeq 6000 platform (Illumina, San Diego, CA) using paired end 150 bp chemistry. Raw data processing followed GATK Best Practices (v4.0), including FastQC QC, Trimmomatic adapter removal, BWA-MEM alignment to hg38 reference, and GATK processing (MergeBamAlignment, MarkDuplicates, BaseRecalibrator, and ApplyBQSR).

**Somatic mutations.** Mutations with > 2% allele frequency were retained for analysis. Mutational burden was calculated per megabase of sequenced exome (38 Mb reference size). Mutational signatures were extracted using maftools (v2.18.0) and compared against COSMIC v2 database. Bootstrap analysis of intervals for exposures was conducted using QP methods in Sigminer R package (v2.3.1). We used SigProfiler (SigProfilerExtractor 1.1.17) to assess the stability of the mutation signature results obtained with maftools. Intratumor heterogeneity was quantified using MATH scores, calculated as the ratio of median absolute deviation to median mutant-allele fraction. Pathway-level mutation impact was assessed using PMAPscore (v0.1.1)<sup>106</sup>.

**Copy number variants.** Copy number identification and tumor purity estimation were conducted via Sequenza (v2.1.9999b1)<sup>51</sup> with default settings. GISTIC2.0 identified SCNAs. Differences of Mutation or CNV change in each BrMS were compared by fisher test. Structural variants were called using Manta at default parameters<sup>55</sup>. To assess both cis- and trans-regulatory effects of SCNAs on mRNA, and protein, Spearman's rank correlation was computed and adjusted by BH method. The results were visualized using the multiOmicsViz R package (v1.24.0). Associations with FDR below 0.05 were considered statistically significant for SCNA-mRNA and SCNA-protein, whether positive or negative.

**Targeted energy metabolomics and therapeutic vulnerabilities**  
**Sample preparation and extraction.** Frozen BrM samples ( $n = 74$ ) were thawed on ice. Approximately 50 mg ( $\pm 2.5$  mg) of each sample was weighed and homogenized with 500  $\mu$ L of pre-cooled ( $-20$  °C) 70% methanol/water (v/v) extraction solvent. Samples were vortexed at 2500 rpm for 3 min and centrifuged at 12,000 rpm for 10 min at 4 °C. The supernatant (300  $\mu$ L) was transferred to a new tube and stored at  $-20$  °C for 30 min, followed by a second centrifugation at 12,000 rpm for 10 min at 4 °C. The final supernatant (200  $\mu$ L) was filtered through a protein precipitation plate prior to LC-MS/MS analysis.

**UPLC-MS/MS analysis.** Chromatographic separation was performed on a Waters ACQUITY H-Class UPLC system equipped with an ACQUITY UPLC BEH Amide column (2.1  $\times$  100 mm, 1.7  $\mu$ m). The mobile phases consisted of water containing 10 mM ammonium acetate and 0.3% ammonium hydroxide (solvent A) and 90% acetonitrile/water (v/v) (solvent B). The gradient elution program was as follows: 0-1.2 min, 95% B; 8 min, 70% B; 9-11 min, 50% B; 11.1-15 min, 95% B. The flow rate was 0.4 mL/min, column temperature was 40 °C, and injection volume was 2  $\mu$ L.

Mass spectrometric detection was conducted using a QTRAP<sup>®</sup> 6500+ LC-MS/MS system (SCIEX) equipped with an electrospray ionization (ESI) source operating in both positive and negative ion modes. The ESI source parameters were set as follows: ion spray voltage, +5500 V (positive mode) and -4500 V (negative mode); source temperature, 550 °C; curtain gas, 35 psi. Scheduled multiple reaction monitoring (MRM) was employed for metabolite quantification, with declustering potential (DP) and collision energy (CE) optimized for each MRM transition. Data acquisition and quantification were performed using Analyst (v1.6.3, SCIEX) and MultiQuant (v3.0.3, SCIEX), respectively.

**Quantification and quality control.** A total of 75 energy metabolism-related metabolites were quantified using an in-house database

(MWDB, Metware). Absolute quantification was achieved using external standard curves prepared at 20 concentration levels (0.01-15,000 ng/mL for most metabolites). Mixed QC samples were prepared by pooling equal volumes of all sample extracts and analyzed 3 QC samples throughout the analytical run. Instrument stability was evaluated by overlapping total ion chromatograms (TICs) of QC samples. The Pearson correlation coefficients among QC samples were > 0.99, and the coefficient of variation (CV) was < 30% for > 80% of detected metabolites.

**Metabolism analysis.** Batch-corrected metabolomic data underwent orthogonal partial least squares discriminant analysis (OPLS-DA). Significant metabolites were defined by VIP > 1 and |LogFC| > 0.58. Pathway enrichment analysis was performed using MetaboAnalyst 6.0<sup>107</sup>. DA scores were calculated to assess pathway-level metabolic alterations<sup>108</sup>. The DA score can indicate the average and gross alterations of all measured metabolites in a pathway. A DA score of 1 indicates all metabolites in a pathway increased in abundance, while a DA score of -1 denotes all metabolites decreased. To investigate metabolic differences between metastatic lesions and matched primary tumors, we performed integrative metabolic pathway enrichment analysis based on differentially expressed genes and differentially expressed proteins from the RNA and proteomic datasets, respectively. KEGG-defined metabolic pathways were used as the reference<sup>99</sup>.

**Master Regulator (MR) analysis for therapeutic prediction.** To enable context-specific inference of regulatory protein activity, we employed VIPER-OncoTreat in combination with ARACNe algorithm. Molecular interaction networks were constructed using ARACNe with 100 bootstrap iterations as previously described<sup>58</sup>. RNA-seq data from BrM-PDO and matched primary/BrM tumors were integrated for network construction. Differential activity analysis between BrMs and matched primary tumors allowed VIPER to identify key Master Regulator (MR) with significantly altered activity. The top 30 most activated and top 30 most inactivated MRs in BrMs were identified. To evaluate therapeutic potential, we employed OncoTreat to identify compounds capable of significantly reversing the activity of MRs constituting the tumor checkpoint module (TCM) that governs metastatic progression<sup>58,59,109</sup>. This analysis identified gamitrinib as a candidate compound with therapeutic potential. The inhibitory effects of compounds candidates were validated in PDOs through IC<sub>50</sub> assays.

**PDO models and drug sensitivity screening.** PDOs were established from surgically resected BrM tissues representing distinct molecular subtypes following previously described protocols<sup>110</sup>. Freshly resected BrM tissues were obtained and mechanically minced into small fragments and enzymatically dissociated in serum-free advanced DMEM/F-12 medium (Gibco) containing 1 mg/mL collagenase IV (Sigma) for 1 h at 37 °C with gentle agitation. The dissociated cell suspension was mixed with Matrigel (BD Biosciences) at a 1:1.5 (v/v) ratio and seeded in 96-well plates (10  $\mu$ L per well). PDOs were cultured in complete organoid medium consisting of advanced DMEM/F-12 supplemented with penicillin/streptomycin (1 $\times$ ), glutamine (1 $\times$ ), B27 supplement (1 $\times$ ), nicotinamide (5 mM), N-acetylcysteine (1.25 mM), A83-01 (500 nM), SB202190 (500 nM), Y-27632 (5  $\mu$ M), Noggin (100 ng/mL), R-spondin 1 (250 ng/mL), FGF2 (5 ng/mL), FGF10 (10 ng/mL), and EGF (5 ng/mL). Culture medium (100  $\mu$ L) was added to each well, and organoids were maintained at 37 °C in a humidified incubator with 5% CO<sub>2</sub>. All PDO cultures were routinely tested for mycoplasma contamination using the MycoAlert Mycoplasma Detection Kit (Lonza) and confirmed to be negative.

To evaluate subtype-specific drug responses, PDOs were subjected to dose-response studies with gamitrinib, everolimus, and

abemaciclib. For drug sensitivity assessment, PDOs were dissociated into clusters (~2000 cells/cluster) and seeded in 384-well plates. Following 48-h culture establishment, cells were exposed to four-fold serial dilutions of each drug: gamitrinib (20, 5, 1.25, 0.3125, 0.078125, 0.01953125  $\mu$ M), everolimus (4, 1, 0.25, 0.0625, 0.015625, 0.00390625  $\mu$ M), and abemaciclib (4, 1, 0.25, 0.0625, 0.015625, 0.00390625  $\mu$ M). Each drug concentration was tested once. Cell viability was evaluated after 72 h using CellTiter-Glo 3D assay (G9681, Promega) according to manufacturer's instructions. Viability measurements were normalized to DMSO-treated controls (100% viability), and IC<sub>50</sub> values were calculated using GraphPad Prism (v10.6.1).

**Computational deconvolution of the tumor immune micro-environment.** RNA-based immune profiling was conducted using MCP counter R package (v1.1)<sup>62</sup> and Cibersort according to LM22 signature<sup>63</sup>. We also inferred and compared the immune cell subtype signature and metabolic signatures, which were collected by Zeng et al.<sup>11</sup>, using single-sample Mann–Whitney–Wilcoxon gene set test (ssMWW-GST)<sup>112</sup>.

**The correlation analysis of immune scores.** To identify potential regulators of antitumor immune activity, we correlated MCP counter CTL scores with proteomics data using Spearman's correlation analysis. GSEA was then performed for Hallmark pathways using signed  $-\log_{10}$  *P*-values. To further evaluate the relationship between CTL scores and EMT activities in BrMs and primary cancers, we applied single-sample GSEA (ssGSEA)<sup>113</sup> using a EMT gene set from HALL-MARKS in MSigDB. We also applied multiple published EMT signatures and performed correlation analyses across four independent datasets in this study.

**mIF staining.** Six BrM lesions were subject to mIF staining. Formalin-fixed paraffin-embedded tissue sections were prepared and stained following the manufacturer's instructions using the Opal Polaris™ 7-Color Manual IHC Kit (Akoya Biosciences). The mIF panel included the following markers: DAPI (Abcam, ab104139), anti-CD4 (Abcam, ab133616, 1:500), anti-CD68 (Abcam, ab213363, 1:500), anti-CD20 (HUABIO, HA721138, 1:3000), anti-panCK (Abcam, ab7753, 1  $\mu$ g/mL), anti-CD8 (Abcam, ab237709, 1:2000), and anti-CD31 (Abcam, ab182981, 1:2000). For signal removal between staining cycles, the petri dish was sandwiched between two broad-spectrum LED light sources for 45 min at 4 °C. After 45 min, sample slides were transferred to a new petri dish with fresh bleaching solution and photobleached for another 45 min at 4 °C.

**Digital spatial profiling.** The processed and normalized DSP data was obtained from Schoenfeld et al.<sup>78</sup>. Correlation analysis between fibronectin and *GZMA/GZMB* expression was performed within the tumor regions (CK-positive areas) of both brain metastasis and primary tumor samples.

### Single-nucleus RNA sequencing

**Nuclei isolation.** Nuclei were isolated from fresh frozen tissue samples using the Boyou® Cell Nucleus Separation Kit. Briefly, tissue samples were homogenized in lysis buffer (LB) supplemented with 1% BSA using a tissue grinder. After incubation on ice for 1–10 min, the lysate was filtered through a 40  $\mu$ m cell strainer and centrifuged at 500 g for 5 min at 4 °C. The pellet was resuspended in LB and layered with PB1, PB2, and PB3 solutions for density gradient centrifugation at 3000 g for 20 min at 4 °C. The nuclei layer was collected and washed with nuclei buffer (NB), passed through a 40  $\mu$ m cell strainer, and centrifuged at 500 g for 5 min at 4 °C. The final nuclei pellet was resuspended in NB. Assess RNA integrity using the Agilent 2100 Bioanalyzer and quantify RNA using the Qubit RNA HS Assay Kit.

**Single-nucleus 3' RNA-seq library preparation.** snRNA-seq libraries were prepared using the 10x Genomics Chromium Next GEM Single Cell 3' Kit v3.1, following the manufacturer's protocol (10x Genomics, 2021). Briefly, nuclei suspensions were loaded onto a Chromium Next GEM Chip G at a concentration of 1000 nuclei/ $\mu$ L. Gel Bead-in-Emulsons (GEMs) were generated using the Chromium Controller. Reverse transcription was performed within the GEMs at 53 °C for 45 min, followed by emulsion breaking and cDNA amplification for 14 cycles. The cDNA was then fragmented, end-repaired, A-tailed, and ligated with adapters. Sample indexing was performed by PCR amplification. Library quality and quantity were assessed using Qubit dsDNA HS Assay Kit and Agilent Bioanalyzer High Sensitivity DNA chip. Libraries were pooled and sequenced on a NovaSeq 6000 system (Illumina) with a recommended configuration of 28 cycles for Read 1, 8 cycles each for i7 and i5 indexes, and 91 cycles for Read 2.

**Single-nucleus sequencing.** Reads were processed using the Cell Ranger 3.0.1 (<https://www.10xgenomics.com/>) pipeline with default and recommended parameters. FASTQs from Illumina sequencing output were aligned to the human genome. The gene-barcode matrix containing barcoded cells and gene expression counts was imported into Seurat R package (v4.1)<sup>114</sup>. Prior to QC filtering, ambient RNA contamination was removed using CellBender (v0.3.0; <https://github.com/broadinstitute/CellBender>) with the remove-background module. CellBender employs a deep generative model to distinguish true cell signals from background RNA commonly present in droplet-based snRNA-seq data. Cluster analysis of single cell count matrices was performed using the Seurat. Doublets were identified and removed using scDbfFinder (v1.16.0; <https://bioconductor.org/packages/scDbfFinder>) on a per-sample basis with default parameters, with expected doublet rates automatically estimated from cell recovery counts.

For QC, we retained cells with 500–10,000 detected genes and 500–60,000 total UMI counts, and excluded cells in which mitochondrial transcripts accounted for  $\geq 15\%$  of total counts. After QC filtering and doublet removal, 135,866 high-quality nuclei from 16 brain metastasis samples were retained for downstream analysis. Per-sample QC metrics including UMI counts, gene counts, mitochondrial percentage, and doublet fractions are provided in Supplementary Data 20.

**Single-nucleus sequencing data analysis.** Normalization and scaling were performed after filtering using the 'NormalizationData' and 'ScaleData' functions with default parameters. Principal components for highly variable genes were calculated using 'RunPCA'. After QC, removal of batch effects, and data integration from 16 samples, cells were used in downstream analysis. Clusters were identified using 'FindClusters' with a 0.5 resolution. UMAP visualized clusters in a reduced 2D space. Cluster markers were identified using 'FindAllMarkers'. The R package Cellchat (v2.1.2) inferred interaction mechanisms among TME components across four subtypes<sup>73</sup>. Functions such as 'compareInteractions', 'netVisual\_heatmap', and 'rankNet' analyzed and compared interaction numbers, strengths, and information flow of signaling pathways or ligand-receptor pairs among tissues. The nichenetr R package (v2.0.0) was applied to predict ligand-receptor interactions mediating cell-cell communication<sup>76</sup>. Single-cell meta-program (MP) scores were calculated using the AddModuleScore function from the Seurat package, with gene sets derived from the original study<sup>79</sup>. Among a total of 41 MPs, MPs (1-23, 25-28) which may be associated with BrMs were retained. We performed consensus clustering to stratify the signatures into distinct subgroups. We used 'AddModuleScore' function in Seurat to calculate the EMT signature score (*VIM*, *FNI*, *SNAI1*, *SNAI2*, *ZEB1*, *ZEB2*, *ITGA5*, *ITGB1*, *CDH2*, *SPARC*, *TAGLN*, *MMP2*, *MMP9*, *ITGA2*, *TGFBI*, *SERPINE1*, *LGALS3*, *ILK*, *ITGB3*, *ITGA6*, *TCF4*) in MTCs and CTL score (*PRF1*, *GZMB*, *GZMH*, *GZMA*, *GZML*, *GZLN*).

*NRG1, CST7, IFNG, TNF*) in CD8 T cells. The mean EMT score of tumor cells and the mean CTL score of CD8 T cells were calculated for each sample, and their association was evaluated using Spearman correlation analysis.

### Spatial transcriptomics

**Visium spatial transcriptomics library preparation.** FFPE samples passing RNA QC (DV200 > 30%) were used for spatial transcriptomic library construction and sequencing. Tissue sections (5  $\mu$ m) were mounted onto Visium Gene Expression slides (10x Genomics), baked at 42 °C for 3 h, and dried in a desiccator at room temperature overnight.

Deparaffinization was performed by incubating slides at 60 °C for 2 h, followed by immersion in xylene and rehydration through an ethanol gradient. H&E staining was subsequently performed using Mayer's hematoxylin (Millipore Sigma), bluing reagent (Dako, Agilent), and alcoholic eosin (Millipore Sigma). Stained slides were scanned under a microscope for spatial alignment. Decrosslinking was then performed using 0.1N HCl and TE Buffer (pH 9.0) to release RNA sequestered by formalin fixation.

Stained slides were incubated with the Human Whole Transcriptome Probe Panel (10x Genomics), which consists of paired probes targeting most genes, with the 5' probe containing Small RNA Read 2S and the 3' probe containing a poly-A sequence. Probe pairs were hybridized to target RNA and subsequently ligated to seal the junctions between them, forming single-stranded ligation products. Following RNase treatment and permeabilization, ligation products were released and transferred to the CytAssist instrument (10x Genomics). The poly-A portion of the ligation products was captured by poly(dT) regions of capture probes pre-coated on the Visium slide, which also contain Illumina Read 1 sequences, spatial barcodes, and unique molecular identifiers (UMIs). Captured probes were extended to generate spatially barcoded ligation products and subsequently released from the slide for Sample Index PCR and final library construction.

Libraries were purified using magnetic beads, and quality was assessed by Qubit fluorometry for concentration and Agilent 2100 Bioanalyzer for fragment size distribution.

**Spatial transcriptomics sequencing.** Visium Spatial Gene Expression libraries consisted of Illumina paired-end sequences flanked with P5 and P7 adapters. The 16-bp spatial barcode and 12-bp UMI were encoded in Read 1, while Read 2S was used to sequence the ligated probe insert. Cluster generation and hybridization of sequencing primers were performed according to the Illumina User Guide, and flow cells were loaded onto Illumina sequencing platforms. Paired-end sequencing was performed with the process controlled by Illumina Data Collection Software with real-time data analysis.

**Spatial transcriptomics sequencing data analysis.** Raw reads were aligned to the human reference genome build GRCh38 (hg38) using the Space Ranger (2.1.1) pipeline with default and recommended parameters. Spatial data visualization and interactive exploration were performed using Loupe Browser (v8.0, 10x Genomics). We first normalized the gene-spot expression matrix using SCTransform by Seurat. Batch effect correction across multiple samples was performed using the Harmony R package (v0.1.1), and data were organized using the SingleCellExperiment R package (v1.16.0). Subsequent unsupervised clustering involved PCA (RunPCA), construction of a neighbor graph (FindNeighbors), and identification of clusters using FindClusters with a resolution parameter set to 0.1. The spot-level clusters were named as molecular niches as previously defined<sup>15</sup>. We utilized RCTD, a robust method for cell type deconvolution based on profiles derived from our snRNA-seq data<sup>77</sup>. The ST object was initialized using the SpatialRNA function, incorporating both the spot-gene count matrix and spatial coordinates. A single-cell reference was

constructed via the Reference function, using the annotated cell-gene matrix from our BrM snRNA-seq dataset. Subsequently, the create.RCTD and run.RCTD functions were applied with the 'full' model to perform deconvolution. Spots with a tumor cell fraction greater than 0.37 were defined as tumor-dominant regions. If a tumor-dominant spot either contained T-cell infiltration or was adjacent to any of spots with T-cell infiltration, it was classified as a T-neighborhood tumor region. The DEGs were calculated using Seurat FindMarkers function. Pathway enrichment was performed using the same method as for bulk transcriptomic data.

**PhenoCycler-Fusion 2.0 multiplexed imaging.** Multiplexed imaging was performed on 23 BrM samples using the PhenoCycler-Fusion 2.0 system using the immune-oncology (IO) panel (Akoya Biosciences), formerly known as CODEX. Pre-conjugated Akoya antibodies were obtained with their respective PhenoCycler-Fusion barcodes, while custom antibody conjugation was performed for remaining antibodies using the antibody conjugation kit (Akoya, Cat# 7000009). Briefly, 50  $\mu$ g of carrier-free antibodies were concentrated using 50 kDa MWCO centrifugal filters, treated with disulfide reduction master mix for 30 min, and conjugated with respective PhenoCycler-Fusion barcodes for 2 h at room temperature. Conjugated antibodies were purified through three buffer exchanges and stored in antibody storage buffer. FFPE tissue sections (5  $\mu$ m) were mounted on charged slides, deparaffinized, and rehydrated through graded ethanol series. Antigen retrieval was performed in 1 $\times$  citrate buffer pH 6.0 using a pressure cooker for 20 min. Following equilibration in staining buffer, antibody cocktails supplemented with blocking reagents were applied and incubated overnight at 4 °C. Post-staining processing included fixation in 1.6% paraformaldehyde, methanol treatment, and final fixative incubation. Imaging was performed on the Phenolmager Fusion microscope using Fusion 1.0.6 software. Raw images underwent automated pre-processing including stitching, registration, and background subtraction using the integrated software pipeline. All antibody information is provided in Supplementary Data 24.

**Multiplex immunofluorescence and PhenoCycler-Fusion data analysis.** Images from multiplex immunofluorescence (mIF) staining and PhenoCycler-Fusion 2.0 were analyzed using QuPath software (v0.5.1), as suggested by Akoya Biosciences. Firstly, we employed a deep learning-based software, StarDist (v0.9.1), to do the cell segmentation. To automate cell type identification, we manually labeled a representative subset of cell types in QuPath. Morphological and staining features were extracted and used to train a built-in random forest classifier, which was subsequently applied to classify all detected cells across the tissue section. The mean staining intensity of each cell was used as a proxy for protein expression, and the centroid coordinates of individual cells were extracted for subsequent spatial analyses. For the PhenoCycler-Fusion dataset, batch effects across different samples were corrected using Harmony integration. Focusing on tumor cells, we applied the Rphenograph R package (v0.99.1) to perform unsupervised clustering, which resulted in the identification of 37 tumor cell clusters. Based on the expression levels of Keratin14, Vimentin, E-cadherin, Pan-Cytokeratin, Keratin8-18, EpCAM, Keratin5, Vimentin, CD44, CD57, these clusters were further grouped and biologically annotated into five major functional subtypes. The spatial analysis was conducted by R package 'imcRtools' R package (v1.5.3) and Squidpy. The distances between cells were calculated using the Euclidean distance between their centroid coordinates. Cell-cell interactions were defined using a 15-micron proximity threshold. To evaluate the spatial interaction patterns between cell types, we used the squidpy.gr.nhood\_enrichment function. This method quantifies the over- or under-representation of each cell type in the neighborhood of another, based on a spatial connectivity graph. Statistical significance is assessed using permutation testing, and results are

returned as an enrichment Z-score matrix indicating pairwise spatial co-enrichment.

**Statistics and reproducibility.** The analysis was conducted using R (v4.2.0). All statistical tests were two-sided. Fisher's exact test was employed for categorical variables. The Wilcoxon rank-sum test and *t* test were used to assess differential expression between two groups, and the Kruskal–Wallis test was used to examine differences among multiple groups. Benjamini–Hochberg FDR correction or Holm's method was applied to adjust *P*-values for multiple testing where indicated. Kaplan–Meier plots with log-rank test were used to depict survival, and multivariate Cox proportional hazards regression models were used to identify variables associated with survival outcomes, adjusting for subtypes, primary tumor site, age, sex, and cohort. No significant differences were observed for age or sex. Spearman correlation analysis was performed to assess correlations. No statistical method was used to predetermine sample size. No data were excluded from the analyses. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

### Reporting summary

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

### Data availability

All newly generated raw sequencing data from this study, including WES, bulk RNA-seq, snRNA-seq, targeted metabolomics, ST, and spatial proteomics have been deposited in the Genome Sequence Archive (GSA) in the National Genomics Data Center, China National Center for Bioinformatics/Beijing Institute of Genomics, Chinese Academy of Sciences. WES and bulk RNA-seq data are available under accession code [HRA010095](#) and [HRA010679](#). snRNA-seq data are available under accession code [HRA010723](#). ST data are available under accession code [HRA015579](#). Due to patient privacy protection requirements and local policy of human genetic resource management, raw sequencing data are available under controlled access. The guidelines for data requests can be found at <https://ngdc.cncb.ac.cn/gsa-human/document>. There are no access restrictions for non-profit research organizations. Data access requests will be promptly reviewed by the Data Access Committee (DAC) within two weeks, notifying GSA for Human of the approval decision. After access approval, raw data files can be downloaded directly from the corresponding FTP directory within three months. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via iProX repository and are publicly available with the dataset identifier [PXD059751](#). Targeted metabolomics data are publicly available in MetaboLights under accession code [MTBLS13554](#). All publicly available datasets used for analysis were retrieved from the following repositories: the European Genome-phenome Archive (EGAS00001003672)<sup>5</sup>, the NGDC GSA-Human repository (HRA004247 and HRA005036)<sup>7</sup>, the NCBI BioProject database (PRJNA681304)<sup>8</sup>, dbGaP (phs002457.v1.p1<sup>25</sup> and phs002639.v1.p1<sup>26</sup>), the NCBI Gene Expression Omnibus (GEO) database (GSE245467<sup>27</sup>, GSE164150<sup>28</sup>, GSE184869<sup>29</sup>, GSE200563<sup>30</sup>, GSE209998<sup>32</sup>, GSE110590<sup>33</sup>, GSE125989<sup>36</sup>, GSE18549<sup>37</sup>, GSE159407<sup>38</sup>), the MET500 cohort<sup>31</sup> from UCSC Xena browser (<https://xenabrowser.net/datapages/>), combined GEO datasets as previously described (PMID: 32591432), and the ArrayExpress database (E-MTAB-8659)<sup>35</sup>. All data necessary to reproduce the analyses reported in this study are available in Source Data file. All processed data generated, including gene expression matrices and analysis outputs, are publicly available in Zenodo repository under <https://doi.org/10.5281/zenodo.17879312>, or can be obtained from the lead corresponding author, Dr. Zhang (gzhang6@me.com), upon request. Source data are provided with this paper.

### References

- Soffiotti, R., Ahluwalia, M., Lin, N. & Rudà, R. Management of brain metastases according to molecular subtypes. *Nat. Rev. Neurol.* **16**, 557–574 (2020).
- Suh, J. H. et al. Current approaches to the management of brain metastases. *Nat. Rev. Clin. Oncol.* **17**, 279–299 (2020).
- Brastianos, P. K. et al. Genomic characterization of brain metastases reveals branched evolution and potential therapeutic targets. *Cancer Discov.* **5**, 1164–1177 (2015).
- Shih, D. J. H. et al. Genomic characterization of human brain metastases identifies drivers of metastatic lung adenocarcinoma. *Nat. Genet.* **52**, 371–377 (2020).
- Fischer, G. M. et al. Molecular profiling reveals unique immune and metabolic features of melanoma brain metastases. *Cancer Discov.* **9**, 628–645 (2019).
- Nguyen, B. et al. Genomic characterization of metastatic patterns from prospective clinical sequencing of 25,000 patients. *Cell* **185**, 563–575.e511 (2022).
- Duan, H. et al. Integrated analyses of multi-omic data derived from paired primary lung cancer and brain metastasis reveal the metabolic vulnerability as a novel therapeutic target. *Genome Med.* **16**, 138 (2024).
- Fukumura, K. et al. Multi-omic molecular profiling reveals potentially targetable abnormalities shared across multiple histologies of brain metastasis. *Acta Neuropathol.* **141**, 303–321 (2021).
- Wolf, D. M. et al. Redefining breast cancer subtypes to guide treatment prioritization and maximize response: Predictive biomarkers across 10 cancer therapies. *Cancer Cell* **40**, 609–623.e606 (2022).
- Verhaak, R. G. et al. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell* **17**, 98–110 (2010).
- Guinney, J. et al. The consensus molecular subtypes of colorectal cancer. *Nat. Med.* **21**, 1350–1356 (2015).
- Soria, J. C. et al. Osimertinib in untreated EGFR-mutated advanced non-small-cell lung cancer. *N. Engl. J. Med.* **378**, 113–125 (2018).
- Solomon, B. J. et al. Post hoc analysis of lorlatinib intracranial efficacy and safety in patients with ALK-positive advanced non-small-cell lung cancer from the phase III CROWN study. *J. Clin. Oncol.* **40**, 3593–3602 (2022).
- Freedman, R. A. et al. TBCRC 022: a phase II trial of neratinib and capecitabine for patients with human epidermal growth factor receptor 2-positive breast cancer and brain metastases. *J. Clin. Oncol.* **37**, 1081–1089 (2019).
- McArthur, G. A. et al. Vemurafenib in metastatic melanoma patients with brain metastases: an open-label, single-arm, phase 2, multicentre study. *Ann. Oncol.* **28**, 634–641 (2017).
- Margolin, K. et al. Ipilimumab in patients with melanoma and brain metastases: an open-label, phase 2 trial. *Lancet Oncol.* **13**, 459–465 (2012).
- Kluger, H. M. et al. Long-term survival of patients with melanoma with active brain metastases treated with pembrolizumab on a phase II Trial. *J. Clin. Oncol.* **37**, 52–60 (2019).
- Tawbi, H. A. et al. Combined nivolumab and ipilimumab in melanoma metastatic to the brain. *N. Engl. J. Med.* **379**, 722–730 (2018).
- Long, G. V. et al. Combination nivolumab and ipilimumab or nivolumab alone in melanoma brain metastases: a multicentre randomised phase 2 study. *Lancet Oncol.* **19**, 672–681 (2018).
- Gadgeel, S. et al. Updated analysis from KEYNOTE-189: pembrolizumab or placebo plus pemetrexed and platinum for previously untreated metastatic nonsquamous non-small-cell lung cancer. *J. Clin. Oncol.* **38**, 1505–1517 (2020).
- Goldberg, S. B. et al. Pembrolizumab for management of patients with NSCLC and brain metastases: long-term results and

- biomarker analysis from a non-randomised, open-label, phase 2 trial. *Lancet Oncol.* **21**, 655–663 (2020).
22. Reck, M. et al. Systemic and intracranial outcomes with first-line nivolumab plus ipilimumab in patients with metastatic NSCLC and baseline brain metastases from CheckMate 227 part 1. *J. Thorac. Oncol.* **18**, 1055–1069 (2023).
  23. Schmid, P. et al. Atezolizumab plus nab-paclitaxel as first-line treatment for unresectable, locally advanced or metastatic triple-negative breast cancer (IMpassion130): updated efficacy results from a randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet Oncol.* **21**, 44–59 (2020).
  24. Habib, N. et al. Div-Seq: Single-nucleus RNA-Seq reveals dynamics of rare adult newborn neurons. *Science* **353**, 925–928 (2016).
  25. Routh, E. D. et al. Comprehensive analysis of the immunogenomics of triple-negative breast cancer brain metastases from LCCC1419. *Front Oncol.* **12**, 818693 (2022).
  26. Wardell, C. P. et al. Genomic and transcriptomic profiling of brain metastases. *Cancers (Basel)* **13**, 5598 (2021).
  27. Mendoza-Valderrey, A. et al. Immunogenomics and spatial proteomic mapping highlight distinct neuro-immune architectures in melanoma vs. non-melanoma-derived brain metastasis. *BJC Rep.* **2**, 38 (2024).
  28. Su, J. et al. Multi-omics analysis of brain metastasis outcomes following craniotomy. *Front. Oncol.* **10**, 615472 (2021).
  29. Cosgrove, N. et al. Mapping molecular subtype specific alterations in breast cancer brain metastases identifies clinically relevant vulnerabilities. *Nat. Commun.* **13**, 514 (2022).
  30. Zhang, Q. et al. The spatial transcriptomic landscape of non-small cell lung cancer brain metastasis. *Nat. Commun.* **13**, 5983 (2022).
  31. Robinson, D. R. et al. Integrative clinical genomics of metastatic cancer. *Nature* **548**, 297–303 (2017).
  32. Garcia-Recio, S. et al. Multiomics in primary and metastatic breast tumors from the AURORA US network finds microenvironment and epigenetic drivers of metastasis. *Nat. Cancer* **4**, 128–147 (2023).
  33. Siegel, M. B. et al. Integrated RNA and DNA sequencing reveals early drivers of metastatic breast cancer. *J. Clin. Invest* **128**, 1371–1383 (2018).
  34. Garcia-Mulero, S. et al. Lung metastases share common immune features regardless of primary tumor origin. *J. Immunother. Cancer* **8**, e000491 (2020).
  35. Pocha, K. et al. Surfactant expression defines an inflamed subtype of lung adenocarcinoma brain metastases that correlates with prolonged survival. *Clin. Cancer Res.* **26**, 2231–2243 (2020).
  36. Iwamoto, T. et al. Distinct gene expression profiles between primary breast cancers and brain metastases from pair-matched samples. *Sci. Rep.* **9**, 13343 (2019).
  37. Hsu, S. D. et al. Use of gene expression signatures to identify origin of primary and therapeutic strategies for patients with advanced solid tumors. *J. Clin. Oncol.* **28**, 10504–10504 (2010).
  38. Rubio-Perez, C. et al. Immune cell profiling of the cerebrospinal fluid enables the characterization of the brain metastasis micro-environment. *Nat. Commun.* **12**, 1503 (2021).
  39. Zhang, Y., Parmigiani, G. & Johnson, W. E. ComBat-seq: batch effect adjustment for RNA-seq count data. *NAR Genom. Bioinform* **2**, lqaa078 (2020).
  40. Wilkerson, M. D. & Hayes, D. N. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics* **26**, 1572–1573 (2010).
  41. Hoshida, Y. Nearest template prediction: a single-sample-based flexible class prediction with confidence assessment. *PLoS One* **5**, e15543 (2010).
  42. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **102**, 15545–15550 (2005).
  43. Liberzon, A. et al. The molecular signatures database (MSigDB) hallmark gene set collection. *Cell Syst.* **1**, 417–425 (2015).
  44. Distler, U. et al. Drift time-specific collision energies enable deep-coverage data-independent acquisition proteomics. *Nat. Methods* **11**, 167–170 (2014).
  45. Meier, F. et al. Online parallel accumulation-serial fragmentation (PASEF) with a novel trapped ion mobility mass spectrometer. *Mol. Cell Proteom.* **17**, 2534–2545 (2018).
  46. Slobodyanyuk, M. et al. Directional integration and pathway enrichment analysis for multi-omics data. *Nat. Commun.* **15**, 5690 (2024).
  47. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinforma.* **9**, 559 (2008).
  48. Yang, Y. et al. Mutational profile evaluates metastatic capacity of Chinese colorectal cancer patients, revealed by whole-exome sequencing. *Genomics* **116**, 110809 (2024).
  49. Huang, R. S. P. et al. Clinicopathologic and genomic landscape of non-small cell lung cancer brain metastases. *Oncologist* **27**, 839–848 (2022).
  50. Huang, R. S. P. et al. Clinicopathologic and genomic landscape of breast carcinoma brain metastases. *Oncologist* **26**, 835–844 (2021).
  51. Favero, F. et al. Sequenza: allele-specific copy number and mutation profiles from tumor sequencing data. *Ann. Oncol.* **26**, 64–70 (2015).
  52. Tate, J. G. et al. COSMIC: the catalogue of somatic mutations in cancer. *Nucleic Acids Res.* **47**, D941–D947 (2019).
  53. Pecori, R., Di Giorgio, S., Paulo Lorenzo, J. & Nina Papavasiliou, F. Functions and consequences of AID/APOBEC-mediated DNA and RNA deamination. *Nat. Rev. Genet.* **23**, 505–518 (2022).
  54. Roberts, S. A. et al. An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. *Nat. Genet.* **45**, 970–976 (2013).
  55. Chen, X. et al. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics* **32**, 1220–1222 (2016).
  56. Iuchi, T. et al. Frequency of brain metastases in non-small-cell lung cancer, and their association with epidermal growth factor receptor mutations. *Int. J. Clin. Oncol.* **20**, 674–679 (2015).
  57. Argelaguet, R. et al. MOFA+: a statistical framework for comprehensive integration of multi-modal single-cell data. *Genome Biol.* **21**, 111 (2020).
  58. Alvarez, M. J. et al. Functional characterization of somatic mutations in cancer using network-based inference of protein activity. *Nat. Genet.* **48**, 838–847 (2016).
  59. Alvarez, M. J. et al. A precision oncology approach to the pharmacological targeting of mechanistic dependencies in neuroendocrine tumors. *Nat. Genet.* **50**, 979–989 (2018).
  60. Le Rhun, E. et al. EANO-ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up of patients with brain metastasis from solid tumours. *Ann. Oncol.* **32**, 1332–1347 (2021).
  61. Thorsson, V. et al. The Immune Landscape of Cancer. *Immunity* **48**, 812–830.e814 (2018).
  62. Becht, E. et al. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol.* **17**, 218 (2016).
  63. Newman, A. M. et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* **12**, 453–457 (2015).
  64. Rooney, M. S., Shukla, S. A., Wu, C. J., Getz, G. & Hacohen, N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell* **160**, 48–61 (2015).
  65. Berghoff, A. S. et al. Density of tumor-infiltrating lymphocytes correlates with extent of brain edema and overall survival time in patients with brain metastases. *Oncoimmunology* **5**, e1057388 (2016).

66. Santos, L. et al. ENPPI induces blood-brain barrier dysfunction and promotes brain metastasis formation in HER2-positive breast cancer. *Neuro Oncol.* **27**, 167–183 (2024).
67. Chu, Y. et al. Pan-cancer T cell atlas links a cellular stress response state to immunotherapy resistance. *Nat. Med.* **29**, 1550–1562 (2023).
68. Ma, S., Ming, Y., Wu, J. & Cui, G. Cellular metabolism regulates the differentiation and function of T-cell subsets. *Cell Mol. Immunol.* **21**, 419–435 (2024).
69. Raud, B., McGuire, P. J., Jones, R. G., Sparwasser, T. & Berod, L. Fatty acid metabolism in CD8(+) T cell memory: Challenging current concepts. *Immunol. Rev.* **283**, 213–231 (2018).
70. Liu, D. et al. Integrative molecular and clinical modeling of clinical outcomes to PD1 blockade in patients with metastatic melanoma. *Nat. Med.* **25**, 1916–1927 (2019).
71. Ravi, A. et al. Genomic and transcriptomic analysis of checkpoint blockade response in advanced non-small cell lung cancer. *Nat. Genet.* **55**, 807–819 (2023).
72. Brandon, M., Baldi, P. & Wallace, D. C. Mitochondrial mutations in cancer. *Oncogene* **25**, 4647–4662 (2006).
73. Jin, S. et al. Inference and analysis of cell-cell communication using CellChat. *Nat. Commun.* **12**, 1088 (2021).
74. Calandra, T. et al. MIF as a glucocorticoid-induced modulator of cytokine production. *Nature* **377**, 68–71 (1995).
75. Gonzalez, H. et al. Cellular architecture of human brain metastases. *Cell* **185**, 729–745.e720 (2022).
76. Browaeys, R., Saelens, W. & Saeys, Y. NicheNet: modeling intercellular communication by linking ligands to target genes. *Nat. Methods* **17**, 159–162 (2020).
77. Cable, D. M. et al. Robust decomposition of cell type mixtures in spatial transcriptomics. *Nat. Biotechnol.* **40**, 517–526 (2022).
78. Schoenfeld, D. A. et al. Immune dysfunction revealed by digital spatial profiling of immuno-oncology markers in progressive stages of renal cell carcinoma and in brain metastases. *J. Immunother. Cancer* **11**, e007240 (2023).
79. Gavish, A. et al. Hallmarks of transcriptional intratumour heterogeneity across a thousand tumours. *Nature* **618**, 598–606 (2023).
80. DeLellis, R. A. & Shin, S. J. Chapter 9 - Immunohistology of Endocrine Tumors. in *Diagnostic Immunohistochemistry (Second Edition)* (ed. Dabbs, D. J.) 261–300 (Churchill Livingstone, 2006).
81. Meier, F., Geyer, P. E., Virreira Winter, S., Cox, J. & Mann, M. BoxCar acquisition method enables single-shot proteomics at a depth of 10,000 proteins in 100 minutes. *Nat. Methods* **15**, 440–448 (2018).
82. Geyer, P. E. et al. Plasma proteome profiling to assess human health and disease. *Cell Syst.* **2**, 185–195 (2016).
83. Gillet, L. C. et al. Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis\*. *Mol. Cell. Proteom.* **11**, O111.016717 (2012).
84. Pentsova, E. I. et al. Evaluating cancer of the central nervous system through next-generation sequencing of cerebrospinal fluid. *J. Clin. Oncol.* **34**, 2404–2415 (2016).
85. Achrol, A. S. et al. Brain metastases. *Nat. Rev. Dis. Prim.* **5**, 5 (2019).
86. Berghoff, A. S. & Preusser, M. The inflammatory microenvironment in brain metastases: potential treatment target?. *Chin. Clin. Oncol.* **4**, 21 (2015).
87. Fidler, I. J., Yano, S., Zhang, R. D., Fujimaki, T. & Bucana, C. D. The seed and soil hypothesis: vascularisation and brain metastases. *Lancet Oncol.* **3**, 53–57 (2002).
88. Massagué, J. & Obenauf, A. C. Metastatic colonization by circulating tumour cells. *Nature* **529**, 298–306 (2016).
89. Vogelstein, B. et al. Cancer genome landscapes. *Science* **339**, 1546–1558 (2013).
90. Alexandrov, L. B. et al. Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).
91. Klemm, F. et al. Interrogation of the microenvironmental landscape in brain tumors reveals disease-specific alterations of immune cells. *Cell* **181**, 1643–1660.e1617 (2020).
92. Biermann, J. et al. Dissecting the treatment-naive ecosystem of human melanoma brain metastasis. *Cell* **185**, 2591–2608.e2530 (2022).
93. Xing, X. et al. Pan-cancer human brain metastases atlas at single-cell resolution. *Cancer Cell* **43**, 1242–1260.e9 (2025).
94. Tagore, S. et al. Single-cell and spatial genomic landscape of non-small cell lung cancer brain metastases. *Nat. Med.* **31**, 1351–1363 (2025).
95. Quail, D. F. & Joyce, J. A. The microenvironmental landscape of brain tumors. *Cancer Cell* **31**, 326–341 (2017).
96. O'Donnell, J. S., Teng, M. W. L. & Smyth, M. J. Cancer immunoeediting and resistance to T cell-based immunotherapy. *Nat. Rev. Clin. Oncol.* **16**, 151–167 (2019).
97. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
98. Contal, C. & O'Quigley, J. An application of changepoint methods in studying the effect of age on survival in breast cancer. *Comput. Stat. Data Anal.* **30**, 253–270 (1999).
99. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics* **16**, 284–287 (2012).
100. Reddy, J. et al. Predicting master transcription factors from pan-cancer expression data. *Sci. Adv.* **7**, eabf6123 (2021).
101. Aibar, S. et al. SCENIC: single-cell regulatory network inference and clustering. *Nat. Methods* **14**, 1083–1086 (2017).
102. Malta, T. M. et al. Machine Learning Identifies Stemness Features Associated With Oncogenic Dedifferentiation. *Cell* **173**, 338–354.e315 (2018).
103. Synapse.org. TCGA PanCanAtlas Data. in <https://www.synapse.org/#/Synapse:syn4976369>.
104. Kerseviute, I. & Gordevicius, J. aPEAR: an R package for autonomous visualization of pathway enrichment networks. *Bioinformatics* **39**, btad672 (2023).
105. Argelaguet, R. et al. Multi-omics factor analysis—a framework for unsupervised integration of multi-omics data sets. *Mol. Syst. Biol.* **14**, e8124 (2018).
106. Li, X. et al. A novel pathway mutation perturbation score predicts the clinical outcomes of immunotherapy. *Brief Bioinform.* **23**, bbac360 (2022).
107. Pang, Z. et al. MetaboAnalyst 6.0: towards a unified platform for metabolomics data processing, analysis and interpretation. *Nucleic Acids Res.* **52**, W398–w406 (2024).
108. Hakimi, A. A. et al. An integrated metabolic atlas of clear cell renal cell carcinoma. *Cancer Cell* **29**, 104–116 (2016).
109. Mundi, P. S. et al. A transcriptome-based precision oncology platform for patient-therapy alignment in a diverse set of treatment-resistant malignancies. *Cancer Discov.* **13**, 1386–1407 (2023).
110. Wang, H. M. et al. Using patient-derived organoids to predict locally advanced or metastatic lung cancer tumor response: A real-world study. *Cell Rep. Med.* **4**, 100911 (2023).
111. Zeng, D. et al. Enhancing immuno-oncology investigations through multidimensional decoding of tumor microenvironment with IOBR 2.0. *Cell Rep. Methods* **4**, 100910 (2024).
112. Garofano, L. et al. Pathway-based classification of glioblastoma uncovers a mitochondrial subtype with therapeutic vulnerabilities. *Nat. Cancer* **2**, 141–156 (2021).
113. Barbie, D. A. et al. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* **462**, 108–112 (2009).
114. Hao, Y. et al. Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat. Biotechnol.* **42**, 293–304 (2024).

115. Kuppe, C. et al. Spatial multi-omic map of human myocardial infarction. *Nature* **608**, 766–777 (2022).

## Acknowledgements

This study was supported by the National Key Research and Development Program of China grant (no.2022YFC3400405 to L.L.), internal funding from Faculty of Dentistry at HKU (No. 109000034, No. 207051060 to G.Z.), Postdoctoral Research Fund of West China Hospital, Sichuan University (2024HXBH159 to Z.Y.), Postdoctoral Fellowship Program of CPSF (GZB20240495 to Z.Y.), the National Natural Science Foundation of China (NSFC 82203574 to S.W., NSFC 82373386 to Y.M., and NSFC 82303805 to H.D.), Science and Technology Support Program of Sichuan Province (2023YFS0126 to S.W.), and 1.3.5 Project for Disciplines of Excellence of West China Hospital (ZYJC21002 and ZYJC21015 to L.L.). We thank Yi Zhang, Yue Li, Linqiao Tang, Wanli Zhang from the Institute of Clinical Pathology of West China Hospital for technical support with histological staining and multiplex immunofluorescence. We thank Accurate International Biotechnology Co. for their assistance with the organoid techniques. We thank Gaojing (Zhejiang Anji) Precision Medicine Technology Co., Ltd. for providing library preparation and sequencing services. We are grateful to all author teams who generously shared their data, and we acknowledge the efforts of the maintainers and curators of public repositories.

## Author contributions

The study was conceptualized, designed, and financially supported by G.Z., Y.M., L.L., W.J., S.W., H.D. and Z.Y.; H.D., G.K.-K.L., K.M.-Y.K., D.Z., X.W., M.T., and H.L. contributed to clinical sample collection and clinical data interpretation; Z.Y., S.W., H.D., and X.W. contributed data analysis, interpretation, and manuscript editing; Y.Y., Y.K., W.H., C.Z., and Y.L. contributed to pathological assessment and the interpretation of stained sections; Y.D. and Y.J. contributed to the snRNA-seq data analysis; Z.C. contributed PDO experiments. All authors participated in manuscript preparation and approved the final version.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-026-68748-y>.

**Correspondence** and requests for materials should be addressed to Gilberto Ka-Kit Leung, Wang Jia, Yonggao Mou, Lunxu Liu or Gao Zhang.

**Peer review information** *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2026, modified publication 2026

<sup>1</sup>Department of Thoracic Surgery, Institute of Thoracic Oncology, West China Hospital, Sichuan University, Chengdu, China. <sup>2</sup>Department of Neurosurgery/Neuro-Oncology, State Key Laboratory of Oncology in South China, Guangdong Provincial Clinical Research Center for Cancer, Sun Yat-sen University Cancer Center, Guangzhou, China. <sup>3</sup>Faculty of Dentistry, The University of Hong Kong, Sai Ying Pun, Hong Kong SAR, China. <sup>4</sup>Department of Neurosurgery, Beijing Tiantan Hospital, Capital Medical University, Beijing, China. <sup>5</sup>Beijing Neurosurgical Institute, Capital Medical University, Beijing, China. <sup>6</sup>Department of Surgery, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Queen Mary Hospital, Pok Fu Lam, Hong Kong SAR, China. <sup>7</sup>State Key Laboratory of Brain and Cognitive Sciences, The University of Hong Kong, Pok Fu Lam, Hong Kong SAR, China. <sup>8</sup>Department of Pathology, State Key Laboratory of Oncology in South China, Guangdong Provincial Clinical Research Center for Cancer, Sun Yat-sen University Cancer Center, Guangzhou, China. <sup>9</sup>Zhongshan School of Medicine, Sun Yat-Sen University, Guangzhou, China. <sup>10</sup>Department of Neurosurgery, West China Hospital, Sichuan University, Chengdu, China. <sup>11</sup>Guangdong Research Center of Organoid Engineering and Technology, Guangzhou, China. <sup>12</sup>These authors contributed equally: Zhenyu Yang, Shiyu Wei, Hao Duan, Xiuqi Wang, Dainan Zhang, Karrie Mei-Yee Kiang, Yulan Deng, Yuanzhong Yang. ✉ e-mail: [gilberto@hku.hk](mailto:gilberto@hku.hk); [noanswear@hotmail.com](mailto:noanswear@hotmail.com); [mouyg@sysucc.org.cn](mailto:mouyg@sysucc.org.cn); [lunxu\\_liu@aliyun.com](mailto:lunxu_liu@aliyun.com); [gzhang6@me.com](mailto:gzhang6@me.com)