Single Nucleus Total RNA Sequencing of Formalin-Fixed Paraffin-Embedded Gliomas

Ziye Xu, Lingchao Chen, Xin Lin, Yuexiao Lyu, Mofei Zhou, Haide Chen, Heng Zhang, Tianyu Zhang, Yu Chen, Yuanzhen Suo, Qian Liang,* Zhiyong Qin,* and Yongcheng Wang**

Gliomas, the predominant form of brain cancer, comprise diverse malignant subtypes with limited curative therapies available. The insufficient understanding of their molecular diversity and evolutionary processes hinders the advancement of new treatments. Technical complexities associated with formalin-fixed paraffin-embedded (FFPE) clinical samples hinder molecular-level analyses of gliomas. Current single-cell RNA sequencing (scRNA-seq) platforms are inadequate for large-scale clinical applications. In this study, automated snRandom-seq is developed, a high-throughput single-nucleus total RNA sequencing platform optimized for archival FFPE samples. This platform integrates automated single-nucleus isolation and droplet barcoding systems with the random primer-based scRNA-seq chemistry, accommodating a broad spectrum of sample types. The automated snRandom-seq is applied to analyze 116 492 single nuclei from 17 FFPE samples of various glioma subtypes, including rare clinical samples and matched primary-recurrent glioblastomas (GBMs). The study provides comprehensive insights into the molecular characteristics of gliomas at the single-cell level. Abundant non-coding RNAs (ncRNAs) with distinct expression profiles across different glioma clusters and uncovered promising recurrence-related targets and pathways in primary-recurrent GBMs are identified. These findings establish automated snRandom-seq as a robust tool for scRNA-seq of FFPE samples, enabling exploration of molecular diversities and tumor evolution. This platform holds significant implications for large-scale integrative and retrospective clinical research.

1. Introduction

Gliomas are the most common form of brain cancer, originating in the glial cells that support neurons. Approximately 30% of brain and central nervous system tumors, and 80% of all malignant brain tumors, are gliomas.[1] Glioblastoma (GBM) accounts for over half of glioma cases.[2,3] GBM remains incurable, with the majority of patients experiencing recurrence within a year.[3] Over the past decades, there have been limited therapeutic options available for malignant gliomas. A main obstacle to developing new drugs for gliomas is the lack of understanding of the molecular diversity and evolution of these tumors.[4]

Exploring gliomas at the molecular level is particularly challenging due to the technical difficulties of analyzing clinical samples, which are primarily obtained and preserved as formalin-fixed paraffin-embedded (FFPE) blocks.[5] While FFPE samples allow for retrospective research with large-scale samples spanning years or even decades, formalin fixation introduces chemical cross-linking and compromises RNA quality,[6] limiting the

---

Z. Xu, X. Lin, Y. Lyu, H. Chen, Y. Chen, Y. Suo, Y. Wang
Department of Laboratory Medicine of The First Affiliated Hospital & Liangzhu Laboratory
Zhejiang University School of Medicine
Hangzhou 310003, China
E-mail: suoyuanzhen@zju.edu.cn; yongcheng@zju.edu.cn

L. Chen, Z. Qin
Department of Neurosurgery
Huashan Hospital
Fudan University
Shanghai 200040, China
E-mail: qinzhiyong@fudan.edu.cn

M. Zhou, H. Zhang, T. Zhang, Q. Liang
M20 Genomics
Hangzhou 311121, China
E-mail: qian.liang@m20genomics.com

Y. Chen
Zhejiang Key Laboratory of Clinical In Vitro Diagnostic Techniques
Hangzhou 310003, China

Y. Suo
Jiangsu Healthy Life Innovation Medical Technology Co., Ltd
Wuxi 214174, China

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/smtd.202301801

DOI: 10.1002/smtd.202301801

study of glioma transcriptional characteristics by single-cell RNA sequencing.

Several emerging single-cell/nuclei RNA sequencing techniques have been developed to address these challenges. However, oligo(dT)-based or probe-based techniques, such as 10X Genomics fixed RNA assay, snPATHO-seq,[7] and snFFPE-seq,[8] exhibit limitations in sensitivity, transcriptome coverage, and biases toward 3′-end or targeted genes. In our prior work, we developed a droplet- and random primer-based single-nucleus RNA sequencing technique (snRandom-seq) for FFPE tissues.[9] Using this technique, single intact nuclei from FFPE tissues are isolated through deparaffinization, rehydration, and nucleus extraction under mild conditions. Total RNAs are captured using random primers for subsequent reverse transcription and synthesis of the second strand by poly(dA) tailing on the first strand complementary DNAs (cDNAs). The cDNAs in a single nucleus are specifically tagged by a microfluidic barcoding platform, followed by amplification and sequencing. snRandom-seq demonstrates much higher RNA coverage compared to state-of-the-art high-throughput single-cell/nuclei RNA sequencing (sc/snRNA-seq) techniques.[10,11] Additionally, snRandom-seq is capable of detecting more non-coding RNAs and nascent RNAs.[9]

Despite its advantages, the manual operations involved in the single nucleus isolation and droplet barcoding of snRandom-seq hinder the throughput for large-scale sequencing of clinical samples. In this study, we address this limitation by automating the procedures for single nucleus isolation and droplet barcoding. We validate the performance of this automated archival single nuclei RNA sequencing platform using matched FFPE/frozen sample pairs and biological replicates. For the first time, we apply this automated snRNA-seq technique to investigate the molecular characteristics of archival FFPE glioma samples. Our study encompasses 116 492 single nuclei from 17 archival FFPE glioma samples, covering various types, grades, and matched primary and recurrent gliomas, providing a comprehensive and retrospective analysis of the molecular landscape of gliomas at the single-cell level.

2. Result

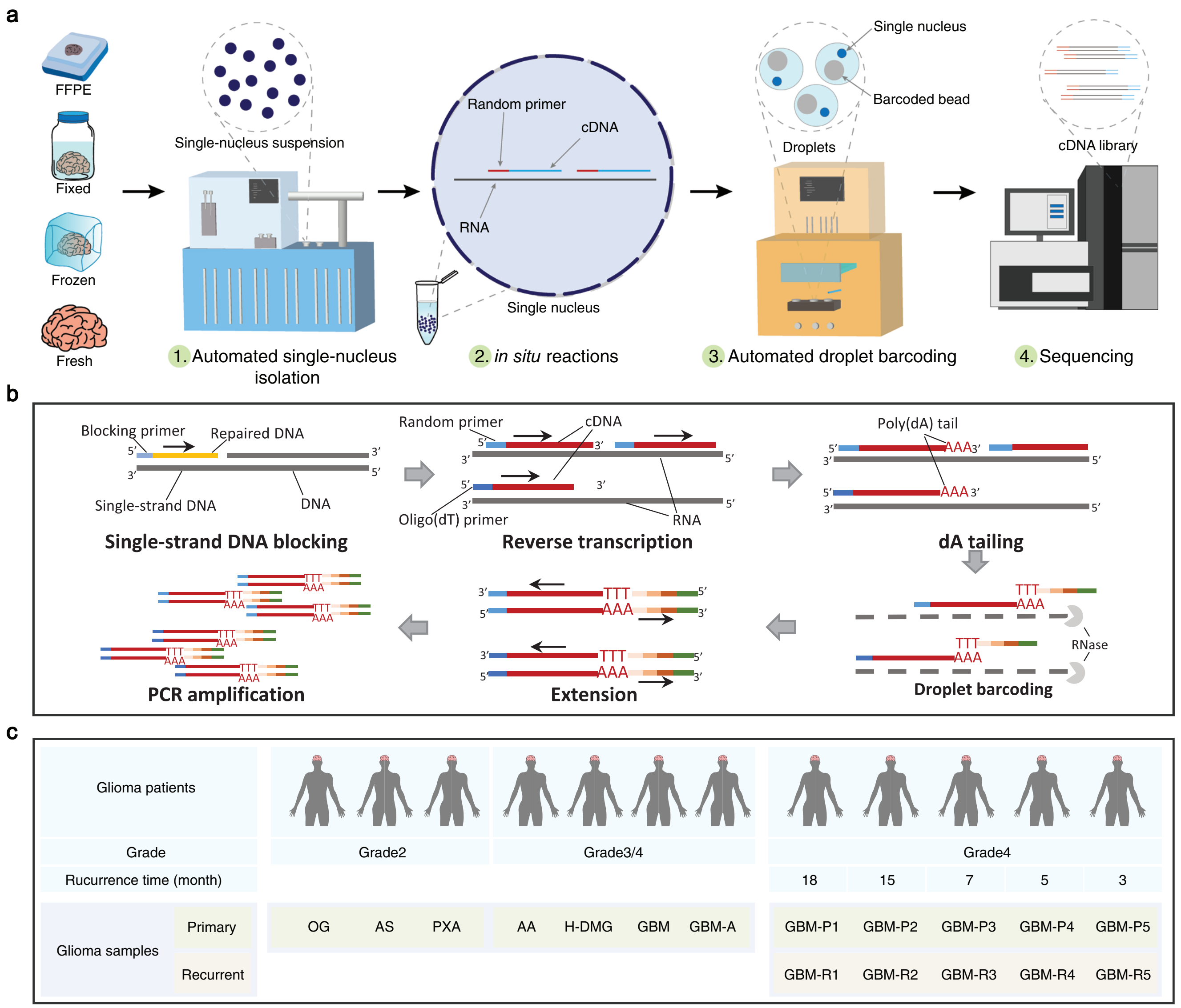
2.1. Design and Validation of an Automated Single Nucleus Total RNA Sequencing Platform

An automated platform for single-nucleus, random-primer-based, total RNA sequencing was designed to enhance the sequencing throughput of clinical FFPE glioma samples. This platform comprises four main subsystems: automated single-nucleus isolation, in situ reactions, automated droplet barcoding, and sequencing (**Figure** 1a). The detailed chemical methods of this automated platform were designed based on our previously published snRandom-seq protocol (Figure 1b). Multiple FFPE glioma samples were processed using the automated snRandom-seq platform (Figure 1c).

The first challenge in large-scale single-nucleus RNA-seq was efficiently isolating clear and intact single nuclei while preserving RNA integrity as much as possible. To address the challenge, we devised an automated single-nucleus isolation system flexible enough to process FFPE, fixed, frozen, and fresh samples. This system comprises a user-friendly interface, a programmable mechanical controller, a temperature controller, and reservoirs for samples and reagents (**Figure** 2a). Accessories such as reagent syringes, pipette tips, grinding rods, reaction tubes, collecting tubes, and cell filters were integrated into the setup (Figure S1a, Supporting Information). Small slices or grains were taken from the regions of interest in the samples and placed into the automated single-nucleus isolation instrument, where they were processed with optimal parameters. The workflow of this system includes four key stages: paraffin dissolution and rehydration (for FFPE samples), lysis, digestion, and filtration (Figure 1b, Supporting Information). The internal structure and flow pathways of the automated single-nucleus isolation are depicted in Figure S1c (Supporting Information). Clean and intact single nuclei, ranging from 5 to 15 µm, were isolated from FFPE, fresh, frozen, and fixed specimens using this automated system (Figure 2b; Figure S1d, Supporting Information). The quality of nuclei isolated by this system is comparable to those obtained through manual operations (Figure S2a,b, Supporting Information). Notably, the lysis and digestive buffers, along with the digestion time, were meticulously optimized to accommodate diverse tissue types in addition to FFPE glioma samples (Table S1, Supporting Information). Solid tissues from parenchymal organs (e.g., liver, muscle, heart) required the strongest lysis and digestive condition. Gastrointestinal tissues (e.g., gastric carcinoma) and immune tissues (e.g., lymphoma) required milder conditions. Upon completion of this procedure, a single-nucleus suspension was obtained, ready for subsequent RNA capture.

RNA molecules within each single nucleus were captured through the multiple annealing of random primers as described in Figure 1b. To address challenges posed by very small samples, minimizing batch effects, or reducing reaction costs, an optional pre-indexing strategy was offered during the reverse transcription step. This strategy involved dividing nuclei into different tubes for reverse transcription, each with specific pre-indexed random primers and subsequently pooling them for further processing. The resulting cDNAs in the reverse transcription step were added with poly(A) tailings, enabling the synthesis of the second DNA strand using barcoded oligo-dT primers.

To achieve precise encapsulation of a single nucleus and a barcoded bead within each droplet, and to alleviate intricate manual procedures, we developed an automated microfluidic droplet barcoding system aided by an artificial intelligence (AI) algorithm (Figure 2c). This system includes various components: a customized pressure pump, a motorized linear slider with tube housing, an imaging unit, a main control panel, and a mechanical stand for microfluidic chip mounting (Figure 2d). The accessories include microfluidic chips and collecting tubes (Figure S3a,b, Supporting Information). Within this system, mixed reagents, single-nucleus suspension, beads, and oil were directly loaded into the microfluidic chip and subsequently propelled into the channel using a homemade pressure pump (Figure 2d). Each individual nucleus with a barcode bead was encapsulated into a droplet, with a size distribution ranging from 80 to 100 µm, similar to previous manual snRandom-seq methods (Figure S2c,d, Supporting Information). Continuous images of droplet generation were captured by a camera and processed by AI image recognition technology integrated into the main control panel (Figure S3c, Supporting Information). By providing microscope images of various droplets for machine learning model



Glioma patients													
Grade		Grade2			Grade3/4				Grade4				
Rucurrence time (month)									18	15	7	5	3
Glioma samples	Primary	OG	AS	PXA	AA	H-DMG	GBM	GBM-A	GBM-P1	GBM-P2	GBM-P3	GBM-P4	GBM-P5
	Recurrent								GBM-R1	GBM-R2	GBM-R3	GBM-R4	GBM-R5

Figure 1. Overview of the automated snRandom-seq platform. a) The flowchart illustrates the complete workflow of the automated snRandom-seq platform. b) A detailed schematic of the chemical methods design used in the automated snRandom-seq platform. c) The table presents details of glioma FFPE samples performed with automated snRandom-seq. OG: oligoastrocytoma, AS: astroglioma, PXA: pleomorphic xanthoastrocytoma, AA: anaplastic astrocytoma, H-DMG: high-grade diffuse midline glioma, GBM: glioblastoma, GBM-A: adjacent tissue of glioblastoma, GBM-P: primary glioblastoma, GBM-R: recurrent glioblastoma.

training, the system can identify droplets that lack beads and those containing one or more beads (Figure S3d, Supporting Information). The initial and final stages of droplet generation were found to be particularly prone to producing unstable droplets. To address this, the main control panel selectively adjusts the motorized linear slider with tube housing based on the number of qualified droplets identified, thereby improving the quality of droplet barcoding in automated snRandom-seq (Figure S3e, Supporting Information). Additionally, the microfluidic runner can sometimes become inadvertently clogged with unexpected impurities during the encapsulation process. When this occurs, the system automatically increases the pressure to flush out the impurities and expels the affected section downstream until the observation window droplets return to normal, after which the system resumes normal operation. For quality control, this automated system droplet barcoding system generates a report that assesses the quality of droplets, including a series of parameters, such as droplet counts, diameters, counts of droplets containing 0, 1, or more than 1 bead, and the rate of qualified droplets (Figure 2d).

The feasibility of the snRandom-seq method was validated in our previous work.[9] In this study, we first compared the performance of the automated and manual versions of snRandom-seq. Results from an FFPE glioma sample adjacent to glioblastoma, analyzed using both automated and manual snRandom-seq, showed a high correlation ($R = 0.97$, $p < 2.2\text{e-}16$) in average gene expression levels (Figure S2e, Supporting Information), as well as a substantial overlap in cell clusters on UMAP plots (Figure S2f, Supporting Information). To further validate

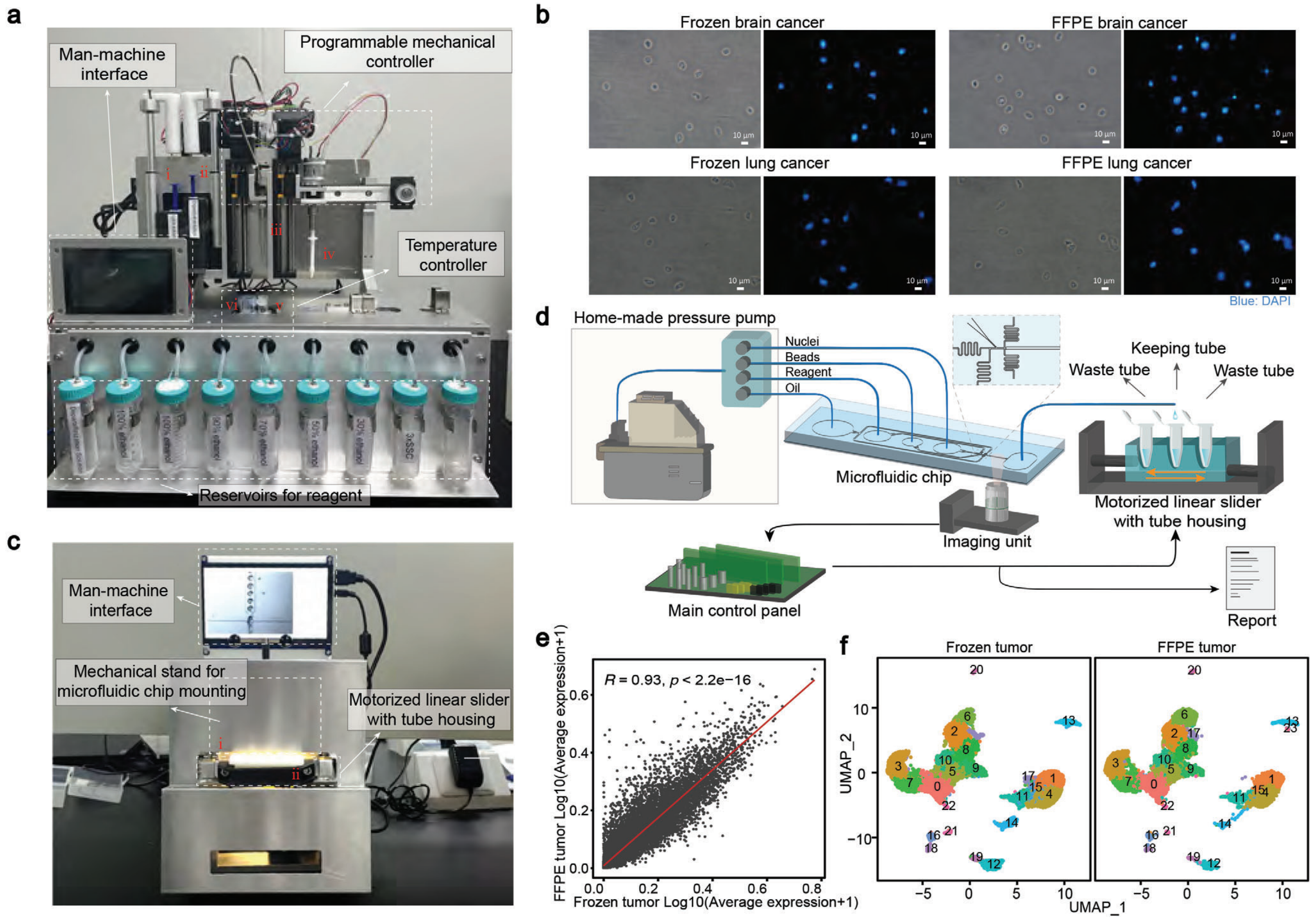


Figure 2. Automated single-nucleus isolation and AI-assistant droplet barcoding systems. a) Photograph depicting the arrangement of components within the automated single-nucleus isolation system of snRandom-seq. The corresponding numerical codes for these system accessories are as follows: i, regent syringe, ii, pipette tip, iii, grinding rod, iv, reaction tube, v, collecting tube, vi, cell filter. b) Microscopical images of nuclei isolated from frozen and FFPE tumor samples from the same brain and lung cancer tissues using automated snRandom-seq, stained with DAPI. Scale bar, 10 μm. c) Photograph depicting the arrangement of various components within the automated droplet barcoding system of automated snRandom-seq. d) Schematic diagram of the automated and AI-assistant droplets barcoding system. e) Dot plot showing the Pearson's correlation coefficient (R) of the average normalized gene expression levels between the frozen and FFPE tumor samples from the same glioblastomas (GBM) tissue. Each dot corresponds to the average normalized expression level of a gene. The red line indicates the linear regression line. The p-value (p) was calculated using a two-sided permutation test. f) Integrated UMAP maps generated from the automated snRandom-seq data of frozen and FFPE tumor samples from the same GBM tissue.

the consistency and reliability of the automated snRandom-seq for gliomas, we applied it to matched FFPE/frozen glioma sample pairs and biological replicates of FFPE glioma samples. High correlations in average gene expression levels were observed in both FFPE/frozen sample pairs ($R = 0.93$, $p < 2.2\text{e-}16$) (Figure 2e) and biological replicates ($R = 0.95$, $p < 2.2\text{e-}16$) (Figure S4a, Supporting Information). UMAP plots of both FFPE/frozen sample pairs (Figure 2f) and biological replicates (Figure S4b, Supporting Information) displayed substantial overlaps in cell clusters. These results provide additional evidence supporting the utility of automated snRandom-seq for glioma samples.

2.2. Integrative Single-Nucleus Atlas of Glioma Subtypes

Previous scRNA-seq studies of gliomas have predominantly focused on single subtype of gliomas, leading to a limited understanding of the comprehensive heterogeneity among various glioma subtypes. The automated snRandom-seq platform for FFPE samples facilitates the exploration of molecular characteristics across various glioma subtypes by FFPE samples. In this study, we performed automated snRandom-seq on FFPE samples of six glioma subtypes, including rare clinical cases, and one adjacent tissue sample (Table S2, Supporting Information). The six glioma subtypes included oligodendroglioma (OG), astrocytoma (AS), pleomorphic xanthoastrocytoma (PXA), glioblastoma (GBM), high-grade diffuse midline gliomas (H-DMG), and anaplastic astrocytoma (AA). Using unsupervised clustering on the integrated automated snRandom-seq dataset of these FFPE samples, nuclei from these distinct samples significantly overlapped within the UMAP plot (**Figure** 3a), separating into 37 clusters based on gene expression patterns (Figure S5a, Supporting Information). According to the expression levels of known markers (Figure S5b, Supporting Information), each cluster was

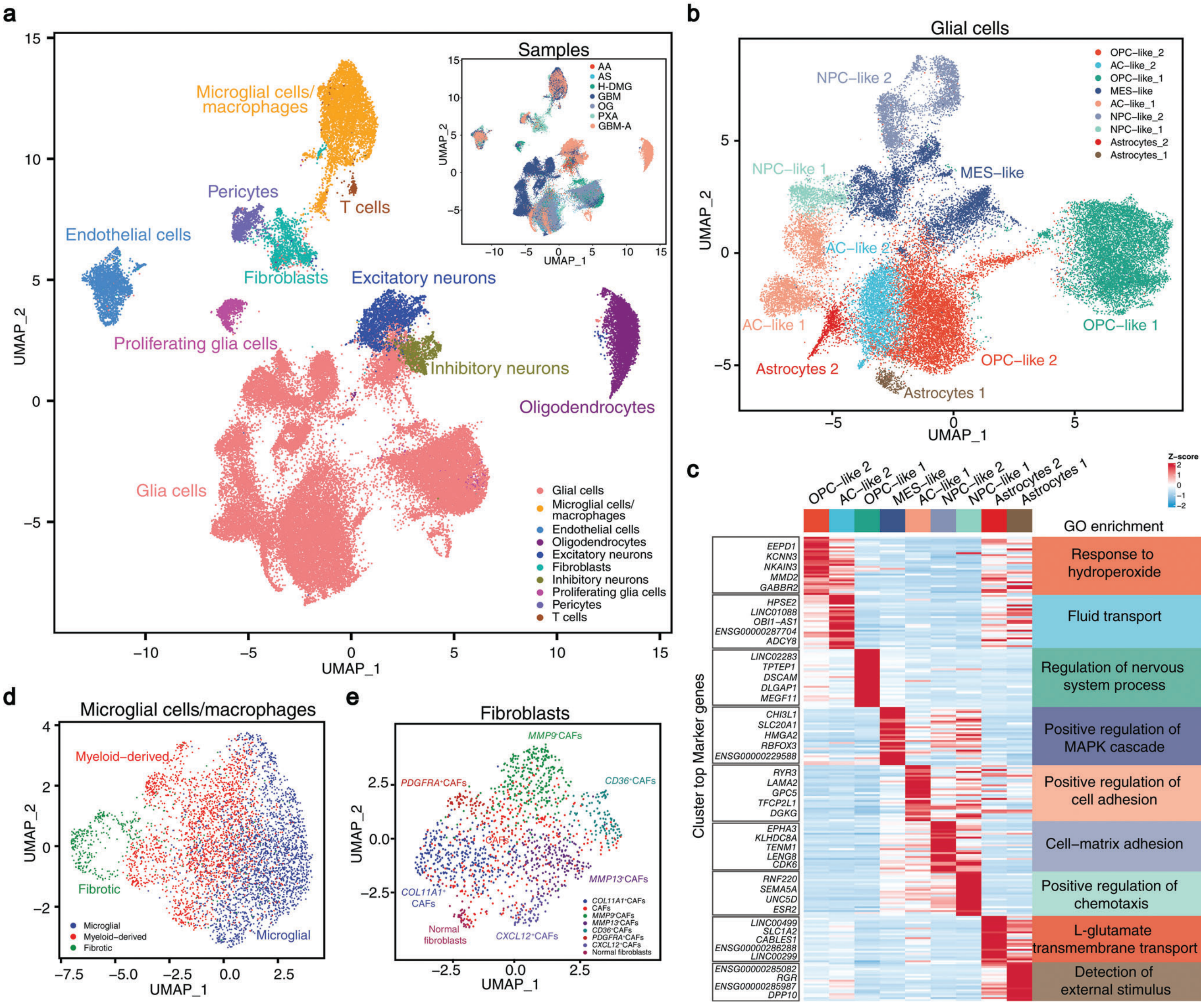


Figure 3. Integrative single-nucleus atlas of glioma subtypes. a) UMAP analysis of nuclei isolated from the seven FFPE samples. The main UMAP plot is colored by identified cell types, with ten cell types identified. The smaller UMAP plot in the upper right corner of **Figure 3**a is colored by samples. OG: oligoastrocytoma, PXA: pleomorphic xanthoastrocytoma, AA: anaplastic astrocytoma, AS: astroglioma, H-DMG: high-grade diffuse midline glioma, GBM: glioblastoma, GBM-A: adjacent tissue of glioblastoma. b) UMAP analysis of the identified glial cells colored by identified subclusters. Nine subclusters of glial cells were identified. OPC: oligodendrocyte-progenitor-like, AC-like: astrocyte-like, MES-like: mesenchymal-like, NPC-like: neural-progenitor-like. c) A heatmap showing the unique top five differentially expressed genes in the nine subclusters of glial cells, ranked by average log2(foldchange). Average gene expression values were scaled and transformed to a scale from −2 to 2. The enriched GO terms of the top 30 differentially expressed genes were shown on the right side of the heatmap. d,e) UMAP analysis of the identified microglial cells/macrophages (d) and fibroblasts (e), colored by identified subclusters. CAFs: cancer-associated fibroblasts.

classified as glial cells, microglial cells/macrophages, T cells, fibroblasts, excitatory neurons, inhibitory neurons, endothelial cells, pericytes, and proliferating glial cells (Figure 3a). Each tumor subtype contained all these identified cell types (Figure S5c, Supporting Information). However, the proportions of cellular composition exhibited notable differences between tumor (GBM) and adjacent tissue (GBM-A), as well as among different glioma subtypes. As expected, glial cells were the most abundant cell type in these gliomas, except in PXA, where fibroblasts constituted 50% of all profiled nuclei. We identified the top five signature genes of each cell type (Figure S5d, Supporting Information). Besides the established cell type markers, such as *MBP* for oligodendrocytes and *COL1A1* for fibroblasts, we discovered potential markers for these cell types in gliomas, such as *SLC4A4* for glial cells. Furthermore, we confirmed the elevated expression of PTPRZ1 in glial cells, previously considered an oncogene implicated in tumor promotion and invasion in gliomas.[12]

We proceeded by isolating glial cells and performing de novo clustering (Figure S6a,b, Supporting Information). The nonmalignant glial cells were categorized as astrocytes-1 and astrocytes-2 (Figure 3b). The malignant glial cells were categorized as OPC-like, NPC-like, AC-like, or MES-like cell states, according to the expression levels of established marker genes (Figure S6c, Supporting Information). Consistent with previous studies,[13] each glioma subtype exhibited a heterogeneous composition of glial cells across these four cell states (Figure S6d, Supporting Information). Compared to other glioma subtypes, GBM exhibited a higher proportion of NPC-like glial cells (NPC-like 1 and NPC-like 2). These particular subpopulations of glial cells expressed distinct gene expression patterns, including both mRNAs and long non-coding RNAs (lncRNAs) (Figure 3c). *CHI3L1*, a gene previously implicated in supporting tumor growth by influencing the state of glioma stem cells,[14] exhibited high expression in the MES-like subcluster. The lncRNA *OBI1-AS1*, recognized as an astrocyte marker with a possible role in glioma recurrence and progression,[15] showed high expression in the AC-like 2 subcluster. We performed Gene Ontology (GO) term enrichment analysis on the top 30 signature genes of these glial cell subpopulations, revealing enrichment in several glioma-related terms (Figure 3c).[16,17] Notably, the signature genes of the astrocytes-2 subcluster were enriched in L-glutamate transmembrane transport, associated with malignant glioma biology.[16] The signature genes of the MES-like subcluster were enriched in positive regulation of MAPK cascade, associated with glioma invasion and metastasis.[17]

Previous studies have identified distinct subpopulations of macrophages and fibroblasts localized around tumor cells as tumor-associated macrophages (TAMs)[18] and cancer-associated fibroblasts (CAFs),[19] holding promise as potential therapeutic targets. We extracted and conducted de novo clustering on the microglial cells/macrophages and fibroblasts obtained from these gliomas, respectively (Figures S7a,b and S8a,b, Supporting Information). We classified the microglial cells/macrophages into three groups: microglial, myeloid-derived, and fibrotic macrophages (Figure 3d), based on their expression patterns of specific cell markers (microglial: *CX3CR1*, *P2RY12*, *P2RY13*, and *SELPLG*, myeloid-derived: *CD163*, *TGFBI*, and *F13A1*, fibrotic: *COL1A1*, *COL1A2*, *COL6A3*, and *COL6A2*) (Figure S7c, Supporting Information). High-grade gliomas (GDM, H-DMG, and AA) exhibited higher percentages of myeloid-derived macrophages (Figure S7d, Supporting Information). Notably, PXA displayed a significant presence of fibrotic macrophages, consistent with the previous PXA case reports describing extensive fibrosis.[20] Furthermore, fibroblasts of these gliomas were further subclustered into distinct categories: Normal fibroblasts, CAFs, *COL11A1*$^+$CAFs, *MMP9*$^+$CAFs, *MMP13*$^+$CAFs, *CD36*$^+$CAFs, *PDGFRA*$^+$CAFs, and *CXCL12*$^+$CAFs (Figure 3e), based on the expression patterns of signature genes (Figure S8c, Supporting Information). PXA exhibited a high percentage of *COL11A1*$^+$CAFs, consistent with its macrophage constitution (Figure S8d, Supporting Information). Subpopulations associated with invasion and metastasis, including *MMP9*$^+$CAFs, *MMP13*$^+$CAFs, and *CXCL12*$^+$CAFs, were mainly present in PXA and H-DMG. Additionally, CellChat analysis of the cell subclusters within the PXA sample highlighted a strong interaction between *PDGFRA*$^+$CAFs and fibrotic macrophages (Figure S8e, Supporting Information).

2.3. Numerous ncRNAs Exhibited Specific Expression within Distinct Glioma Clusters

Multiple studies have indicated that non-coding RNAs (ncRNAs) play critical roles in various biological processes involved in glioma initiation and progression.[21] Most mainstream scRNA-seq studies rely on the prevalent poly(A)-based RNA capture strategy, which targets only polyadenylated mRNAs. This strategy inherently limits the capture of a comprehensive transcriptome, notably missing numerous crucial regulatory RNAs that lack poly(A)-tails. In this study, the automated snRandom-seq platform demonstrated its proficiency in detecting a lot of ncRNAs, including regulatory ncRNAs (lncRNAs and miRNAs) and constitutive ncRNAs (snRNAs and snoRNAs) across all FFPE glioma samples (Figure S9a, Supporting Information). The lncRNAs detected by automated snRandom-seq included over 6500 unannotated and over 3500 annotated lncRNAs (**Figure 4**a). To further characterize these ncRNAs within gliomas, we extracted the expression matrix of ncRNAs from the merged datasets of these FFPE samples and performed a finding-marker analysis to identify signature ncRNAs across various cell types (Figure 4b). All these signature ncRNAs, including *MIR9-1HG* (in glial cells and proliferating glia), *OBI1-AS1*, and *LINC01088* (in glial cells), *LINC01572* (in proliferating glia), *LINC01608*, *LINC00844*, *LINC00639*, and *LINC01170* (in oligodendrocytes), *LINC01374*, *MIR646HG*, and *DLEU1* (in macrophages), *MIR3667HG* (in macrophages and T cells), *LINC00861* (in T cells), *LINC01250* and *MIR124-1HG* (in excitatory neurons and inhibitory neurons), *JARID2-DT* (in excitatory neurons), *DLX6-AS1* (in inhibitory neurons), *TEX41* and *TTTY14* (in fibroblasts), and *CARMN* (in pericytes), were found to be lncRNA genes. These lncRNAs may hold signature roles in identifying specific glioma cell types.

We proceeded to focus on the expression matrix of ncRNAs within glial cells and identified signature ncRNAs among the distinct glial cell subpopulations (Figure 4c). Within this set of ncRNAs, we observed several previously implicated as promoters of glioma progression, such as *LINC02283*,[22] *LINC01088*,[23] *LINC00689*,[24] and *NEAT1*.[25] Particularly, *NEAT1* exhibited high expression in the malignant subclusters mainly found in GBM (MES-like, AC-like 1, NPC-like 1, and NPC-like 2). Consistently, patients with GBM exhibiting high NEAT1 expression displayed reduced survival times (Overall Survival: $P = 0.085$, Disease Free Survival: $P = 0.045$) (Figure S9b,c, Supporting Information). Furthermore, several protein-coding genes associated with these ncRNAs, such as the neighboring coding gene (*SLC7A11*) of signature lncRNA *LINC00499* and the gene (*EMX2*) on the opposite strand transcript of signature lncRNA *EMX2OS*, have been previously reported to be associated with glioma.[26,27] Moreover, we collected information about RNA-RNA/Protein interactions, cancer function states, and related diseases of these signature ncRNAs (Figure 4c). Notably, the signature ncRNAs of AC-like 2 (*LINC01088*, *OBI1-AS1*, *POT1-AS1*, and *LINC00836*) were enriched in the interaction with protein QKI, which plays a regulatory role in glioma stem cell (GSC) stemness.[28] Similarly, the signature ncRNAs of OPC-like 1 (*LINC00689*, and *KCNQ10T1*) and NPC-like 1 clusters (*NEAT1* and *APCDD1L-DT*) were enriched in the interaction with RNAs hsa-miR-346 and hsa-miR-449b-5p, respectively, both of which regulate glioma growth.[29,30]

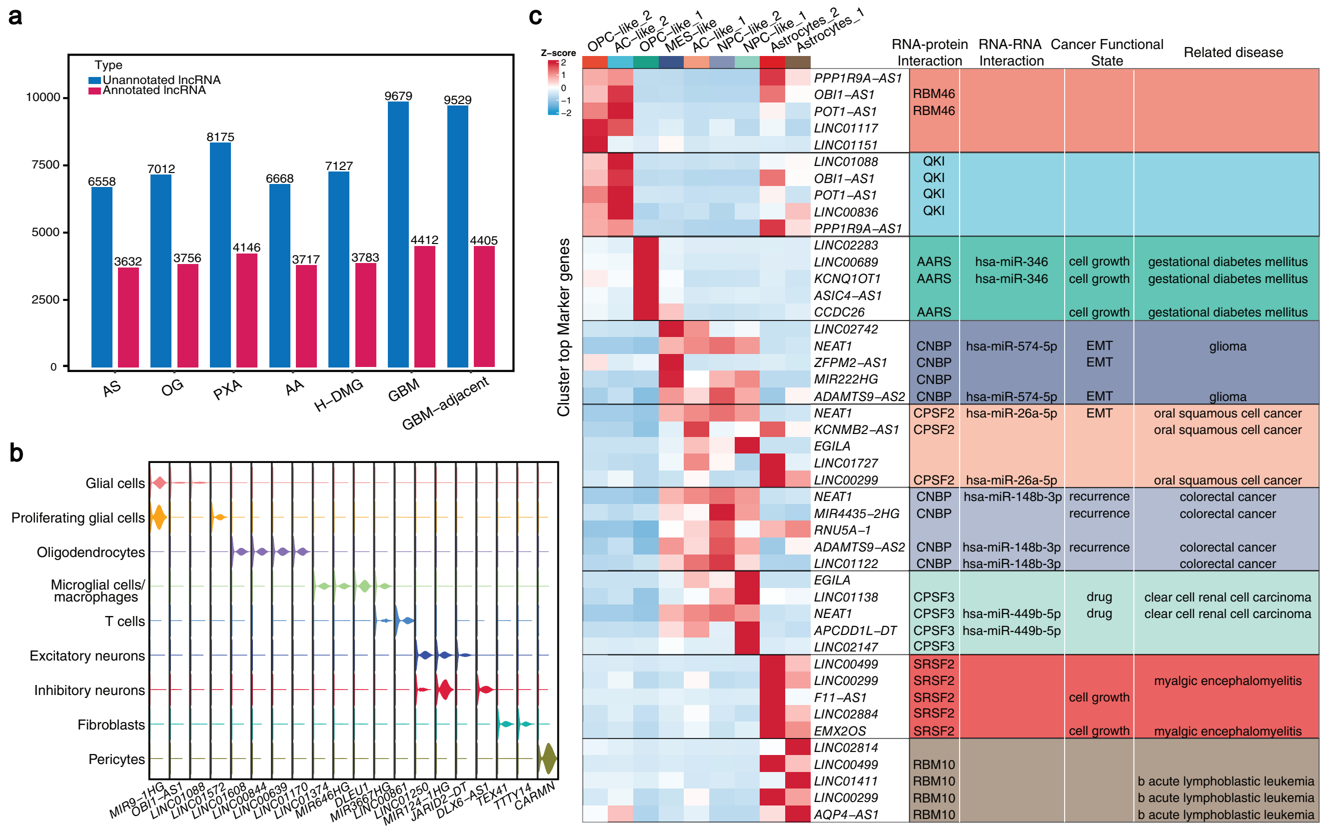


Figure 4. Numerous ncRNAs exhibited specific expression within distinct glioma clusters. a) A bar graph depicting the numbers of unannotated and annotated lncRNAs detected in the seven FFPE samples by automated snRandom-seq. b) Violin plot illustrating the specifically expressed ncRNAs in different cell types of gliomas. c) A heatmap showing the top five signature ncRNAs in the nine subclusters of glial cells, ranked by average log2 (foldchange). Average gene expression values were scaled and transformed to a range from −2 to 2. On the right side of the heatmap, information regarding the interacted RNA, protein, cancer function state, and related diseases of these top 5 signature ncRNAs were provided. This information was collected from LncSEA2.0 (http://bio.liclab.net/LncSEA/index.php).

These signature ncRNAs were enriched in various cancer function states. For example, the signature ncRNAs of OPC-like 1 (*LINC00689*, *KCNQ1OT1*, and *CCDC26*) and Astrocytes 2 clusters (*F11-AS1* and *EMX2OS*) were enriched in cell growth. The signature ncRNAs of MES-like (*NEAT1*, *ZFPM2-AS1*, and *ADAMTS9-AS2*) were enriched in epithelial-mesenchymal transition (EMT), consistent with the mesenchymal-related mRNA gene expression signature of MES-like subcluster. These signature ncRNAs were previously reported to be related to various diseases. Notably, the specific-expressed ncRNAs of MES-like (*NEAT1* and *ADAMTS9-AS2*) are related to glioma. Accordingly, these signature ncRNAs detected by automated snRandom-seq might exert diverse functions at different stages of glial cells and could serve as therapeutic targets for gliomas.

2.4. Retrospective Atlas of Matched Primary-Recurrent Glioblastomas

Conventional scRNA-seq technologies encounter limitations when attempting to utilize matched primary-recurrent samples from the same patients for longitudinal clinical studies.[31] This limitation can potentially result in the omission of crucial information regarding tumor progression. Automated snRandom-seq demonstrated high applicability on archival samples, especially FFPE samples. We applied automated snRandom-seq to five pairs of matched primary-recurrent FFPE samples from GBM patients with different recurrence times (Patient G7: 18 months, Patient G8: 15 months, Patient G9: 7 months, Patient G10: 5 months, Patient G11: 3 months) (Table S2, Supporting Information). Following standard scRNA-seq analysis on the integrated automated snRandom-seq data, we constructed a retrospective atlas of GBMs. Nuclei from primary and recurrent FFPE samples overlapped well on the UMAP plots (Figure S10a,b, Supporting Information) and clustered into distinct cell clusters (Figure S10c, Supporting Information). Main glioma cell types, including glial cells, microglial cells/macrophages, proliferative glia, oligodendrocytes, endothelial cells, pericytes, fibroblasts, and T cells, were identified in all primary and recurrent samples using known markers (**Figure 5**a; Figure S10d, Supporting Information). Notably, the proportion of oligodendrocytes was significantly higher in recurrent sample R1, which had the longest recurrence time (18 months) (Figure S10e, Supporting Information). An increase in macrophages and T-cell infiltration is often observed in recurrent GBM.[32,33] We found increased proportions of immune cells, including microglial cells/macrophages

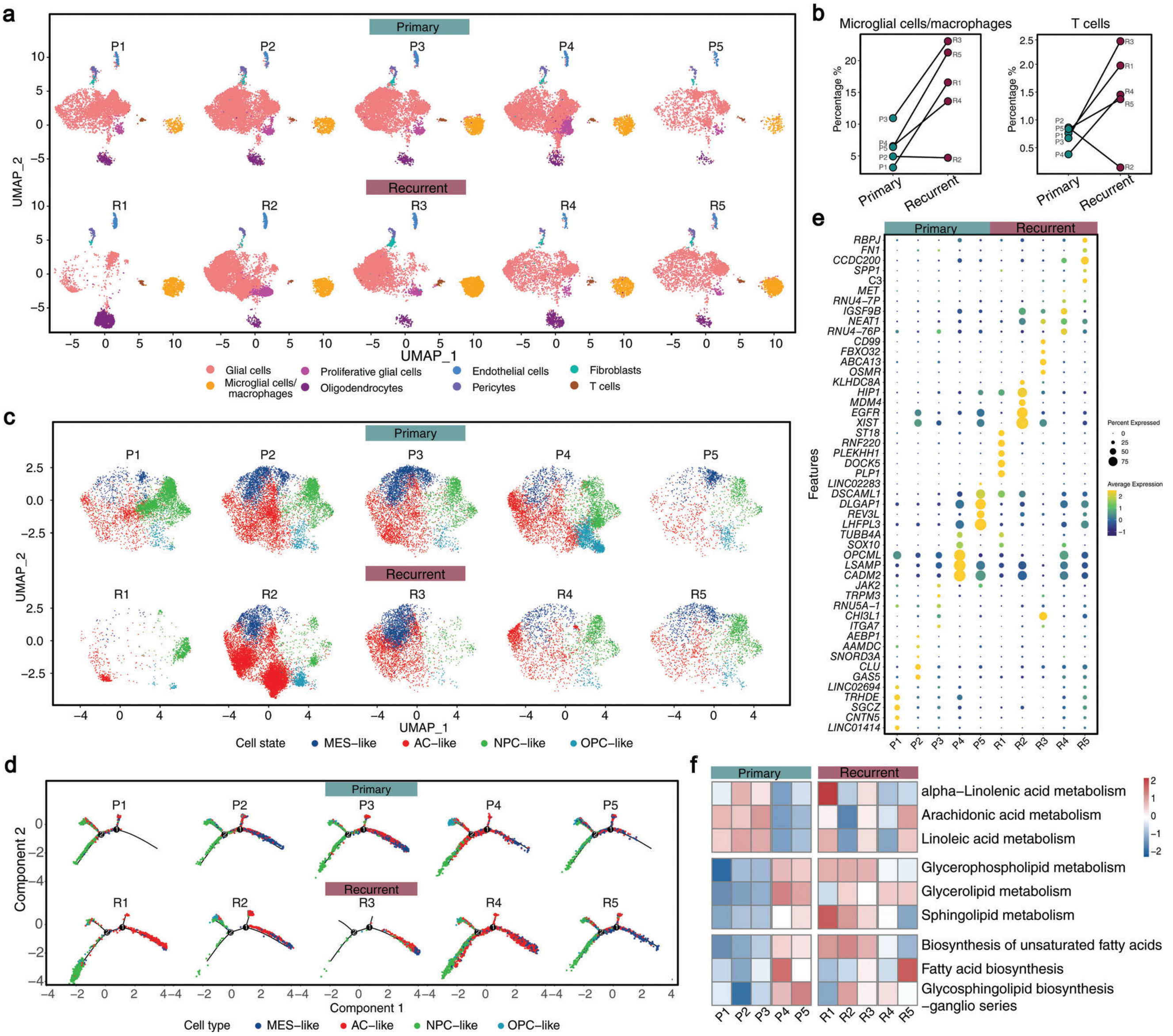


Figure 5. Retrospective atlas of matched primary-recurrent glioblastomas. a) Separated UMAP plots of nuclei from five pairs of matched FFPE samples of primary-recurrent GBM cases. UMAP plots are colored by identified cell types. P1-P5: primary GBM sample 1–5. R1-R5: recurrent GBM sample 1–5. b) The proportions of microglial cells/macrophages and T cells in primary and recurrent samples. Left panel: microglial cells/macrophages, right panel: T cells. c) Separated UMAP plots of glial cells from primary and recurrent GBM samples. UMAP plots are colored by glial cell states. d) Separated trajectories of glial cells from primary and recurrent GBM samples generated by monocle analysis and colored by glial cell states. e) Dot plot showing the unique top five signature genes of OPC-like glial cells among different primary and recurrent GBM samples. f) Heatmap showing the expression scores of lipid metabolism pathways in OPC-like glial cells of different primary and recurrent GBM samples. Source data are provided as a Source Data file.

and T cells, in the recurrent group, except for R2, which had a relatively long recurrence time (15 months) (Figure 5b). Differences in the proportions of oligodendrocyte and immune cells in these recurrent samples might be associated with the recurrence time and individual patient differences. The proportions of microglial cells/macrophages in primary samples with relatively short recurrence time (7, 5, and 3 months) were higher than in primary samples with longer recurrence time (18 and 15 months) (Figure S10e, Supporting Information), suggesting a correlation between microglial cells/macrophages proportion and recurrence time.

Next, we performed de novo unsupervised clustering of glial cells (Figure S11a–c, Supporting Information). Four main glial cell states, including MEC-like, AC-like, OPC-like, and NPC-like clusters, were identified according to previously identified highly expressed genes (Figure 5c; Figure S11d, Supporting Information).[13] The proportions of these four glial subpopulations varied among individual patients (Figure S11e, Supporting Information). Group preference analysis showed that NPC-like and OPC-like clusters were more concentrated in primary samples, while the AC-like cluster was more concentrated in recurrent samples (Figure S11f, Supporting Information). We reconstruct the pseudo-temporal trajectory inference of all glial cells from primary and recurrent GBM using monocle analysis (Figure S11g, Supporting Information). The glial cells trajectory yielded five developmental hierarchies (State 1–5), with the NPC-like cluster mainly located at the beginning of the cell evolution map and the AC-like and MES-like clusters mainly at the endpoint (Figure 5d; Figure S11h, Supporting Information), consistent with the previously reported transition direction along the NPC/OPC-AC-MES axis.[34] Notably, the OPC-like cluster was mainly located at a branch of the starting point (state 5), especially in the primary samples P4 and P5, which had relatively shorter recurrence times. Previous studies have shown that OPC-like cells exhibit greater proliferation and tumor-propagating potential.[35] Combined with the higher proportion of OPC-like cluster in P4 and P5 (Figure S11e, Supporting Information), we speculate that OPC-like glial cells may be associated with GBM recurrence and warrant further therapeutic attention. We identified signature genes of OPC-like cells in these primary and recurrent samples and discovered several glioma-associated genes (*LHFPL3*,[36] *DLGAP1*,[37] and *LINC02283*)[22] highly expressed in the primary samples P4 and P5 (Figure 5e). The oncogenic lncRNA *LINC02283* was also identified as a specifically expressed ncRNA among the glial subpopulations in the integrative single-nucleus atlas of glioma subtypes (Figure 4c), suggesting *LINC02283* may be a critical regulator in glioma biology. Lipid metabolism is reported to be significantly dysregulated in gliomas.[38] We evaluated scores of lipid metabolic pathways in OPC-like glial cells from different samples and found these pathways enhanced in recurrent samples (Figure 5f). Interestingly, unsaturated fatty acid (alpha-linolenic acid, arachidonic acid, and linoleic acid) metabolism was depleted in the primary P4 and P5, which had shorter recurrence times. Glycerophospholipid, glycerolipid, and sphingolipid metabolism were stronger in P4 and P5 than in primary samples with longer recurrence times. Biosynthesis pathways, including biosynthesis of unsaturated fatty acids, fatty acid biosynthesis, and glycosphingolipid biosynthesis-ganglio series, were also stronger in P4 and P5. These results suggest that the accumulation of unsaturated fatty acids in OPC-like glial cells may be positively associated with GBM recurrence.

3. Discussion

Gliomas are characterized by their intrinsic heterogeneity and tendency for recurrence. The advancements in single-cell RNA sequencing (scRNA-seq) technologies have empowered researchers to scrutinize the intricacies of glioma subpopulations, unraveling their composition, functions, and interactions.[19,39–42] In this study, we introduced an automated platform designed for high-throughput single-nucleus total RNAs sequencing of archival FFPE samples. Previous studies often rely on fresh or frozen samples, which can restrict specimen collection from routine clinical archives for investigating glioma subtypes and recurrence. Moreover, previous studies commonly employ oligo(dT)-based scRNA-seq platforms, such as 10X Genomics, which have limitations in transcriptome coverage and biases toward the 3′-end. Consequently, their findings primarily stem from downstream analysis of mRNA expression profiles. In this study, we developed automated snRandom-seq and applied it to diverse glioma subtypes, as well as matched primary-recurrent GBMs. Although our study shares similarities with previous scRNA-seq studies in exploring glioma heterogeneity and cellular composition, it extends these findings by comprehensively examining multiple glioma subtypes and uncovering subtype-specific variations in cellular composition and functional characteristics. Using automated snRandom-seq, our study extends this exploration by detecting a broader range of ncRNAs and identifying signature ncRNAs associated with specific glioma cell types and subpopulations. Furthermore, the longitudinal analysis in this study, using matched primary and recurrent GBM samples, elucidates dynamic changes in cellular composition and molecular pathways associated with tumor recurrence.

A recent advancement introduced a high-throughput single-cell DNA-seq method designed specifically for archival FFPE samples,[5] recognizing the substantial value of clinical FFPE samples in medical research and healthcare. Over the years, millions of FFPE samples have been preserved, often paired with detailed pathological and clinical documentation, making them readily accessible for the study of virtually any disease.[43] It is both important and feasible to evaluate the generalizability of the biological findings presented in this study and to further explore glioma progression in larger cohorts of FFPE glioma samples. Meanwhile, it is essential to conduct molecular biology experiments on more extensive collections of FFPE glioma samples to further validate the molecular characteristics and potential therapeutic targets of glioma, such as the lncRNA *LINC02283*.

The optimized single-nucleus isolation system and random primer-based RNA capture strategy within automated snRandom-seq render it highly versatile, accommodating various sample types, including fresh, frozen, fixed, and FFPE tissues. We are actively exploring opportunities to broaden the applicability of automated snRandom-seq to encompass samples such as blood, saliva, feces, and even microbe, thereby extending diverse clinical application scenarios. This rapid cancer data accumulation demands an inevitable trend of experimental instruments automating. More advanced technologies, such as robotics, artificial intelligence, and machine learning, are integrating into experimental processes. We are committed to the development of a fully automated, sample-in-result-out scRNA-seq platform, aiming to facilitate its utilization in screening, diagnosis, treatment monitoring, and prognosis evaluation across a wide spectrum of conditions.

Our application of automated snRandom-seq to over one hundred clinical cancer FFPE tissues has generated scRNA-seq datasets encompassing total RNA information. We are extending the application of automated snRandom-seq to more types of human cancers. Integration of these scRNA-seq datasets with clinical and basic research data is also currently underway. Additionally, the data preprocessing and traditional analysis of scRNA-seq data are time-consuming, energy-draining, and easy to introduce human error. Many researchers have reported automated pipelines for comparative analysis of scRNA-seq datasets,[44,45] which can be integrated into automated snRandom-seq platform to prompt the widespread applications of this large-scale and comprehensive scRNA-seq database across a wide range of scientific disciplines.

In recent years, the concept of “big data” in clinical settings has evolved significantly, driven by breakthroughs in high-throughput technologies. Alongside this evolution, innovative automation technologies, such as machine learning and microfluidics, have increasingly been applied in the biomedical field, offering promising alternatives to traditional models.[46–48] The rising demand for these automation technologies stems from their ability to deliver more precise, efficient, and cost-effective solutions. A fully automated and integrated high-throughput single cell/nucleus total RNA sequencing platform, which encompasses a kit, instrument, and software, combining the utilization of databases, holds substantial potential to become an essential clinical tool for cancer diagnosis and treatment.

4. Experimental Section

Experimental Model: The collection of human samples and research conducted in this study was approved by the Ethics Committee of Huashan Hospital, Fudan University (approval numbers: KY2022-762) and the Research Ethics Committee of the First Affiliated Hospital, Zhejiang University School of Medicine (approval numbers: IIT20220893A). Clinical FFPE and frozen glioma samples were provided by Huashan Hospital, Fudan University. Informed written consents of all participants for specimen collection and further analysis were obtained. The other samples were provided by the First Affiliated Hospital, Zhejiang University School of Medicine. This study is compliant with the Guidance of the Ministry of Science and Technology (MOST) for the Review and Approval of Human Genetic Resources (approval numbers: 2023BAT0055).

Automated Single-Nucleus Isolation: The setup and accessories of the automated single-nucleus isolation system are shown in Figures 2a and S1a (Supporting Information). The setup includes a man-machine interface, a programmable mechanical controller, a temperature controller, and reservoirs for samples and reagents. The accompanying accessories include syringes for reagents, pipette tips, grinding rods, reaction and collection tubes, and cell filters. For FFPE samples, dewaxing and rehydration reagents were introduced into the tube and incubated with vibration, followed by sequential removal. To ensure complete paraffin removal, the dewaxing step was repeated once or twice, as dictated by the specific sample requirements. The tissue was quickly and completely grinded in the presence of lysis buffer under a low temperature (≈4 °C). Then, tissue was dissociated by digestive enzymes (Collagenase under ≈37 °C, Protein K under ≈50 °C). The resulting mixture of dissociated tissue and buffer was filtered, and the single-nucleus suspension was collected. The detailed internal structure and the flow directions of the automated single-nucleus isolation system are illustrated in Figure 2c. The dewaxing and rehydration regents were sucked and injected into the injector tube in turn by the syringe device. All the waste liquid was siphoned off by the pipette tip at the end of each reaction and stored in the waste liquid tank. The lysis buffer was injected by the syringe device into the injector tube. After lysis and digestion, the mixture was transferred to the cell filter above the collecting tube by the pipette tip.

The lysis buffer, digestive enzyme, digestive time of various tissue types are provided in Table S1 (Supporting Information). The high lysis buffer contained 1X PBS buffer, 0.2% Nonidet(R)P-40 (NP-40, Sangon Biotech, Cat # A600385), and 1 U μL^{-1} RNase Inhibitor. The medium lysis buffer contained 1X PBS buffer, 0.1% TritonX-100 (Sangon Biotech, Cat # A600198), and 1 U μL^{-1} RNase Inhibitor (Yeasen Biotechnology, Cat # 10610ES03). The low lysis buffer contained 1X PBS buffer, 0.1% Tween-20 (Sangon Biotech, Cat # A600560), and 1 U μL^{-1} RNase Inhibitor. Digestive enzymes, including 1 mg mL^{-1} Proteinase K (Sangon Biotech, Cat # A610451) or 1 mg mL^{-1} Collagenase I (Gibco, Cat # 17 100 017) were used.

In Situ DNA Block: The single nucleus suspension was assessed and quantified through DAPI staining under a fluorescent microscope. Block primers, in accordance with the sequence provided in the previous work,[9] were ordered from Sangon Biotech company (China). The reaction mixture was prepared as follows: 100 000–1 000 000 nuclei in 25.5 µL of PBS, 5 µL of 10 µм block primers, 2 µL of DNA Polymerase (M20 Genomics, Cat # R20123124), 10 µL of 5X DNA polymerization buffer, 5 µL of 100 mм dNTP, 2.5 µL of RNase Inhibitor. This mixture was incubated at 37 °C for 30 min. Then, nuclei were subjected to three washes with PBST (1X PBS with 0.05% T-ween 20) to eliminate any residual primers and reagents.

In Situ Reverse Transcription: Following the in situ DNA block, in situ reverse transcription was proceeded. Random primers were ordered from Sangon Biotech company (China) according to the sequence provided in the previous work.[9] The reaction mixture was composed of 100 000–1 000 000 nuclei in 27.5 µL of PBS, 5 µL of 10 µм random primers, 2.5 µL of Reverse Transcriptase (M20 Genomics, Cat # R20123124), 10 µL of 5X reverse transcription buffer, 2.5 µL of 100 mм dNTP, 2.5 µL of RNase Inhibitor. The reaction mix was incubated with twelve cycles of multiple annealing ramping from 8 °C to 42 °C and 30 min at 42 °C on a thermocycler. Then, nuclei were washed with PBST three times to wash away the residual random primers and regents.

dA Tailing: dA tailing was performed after in situ reverse transcription. The following reaction mix was prepared: 100 000–1 000 000 nuclei in 39 µL PBS, 5 µL 10X TdT reaction buffer, 0.5 µL TdT enzyme (NEB, Cat # M0315S), 0.5 µL 100 mм dATP (NEB, Cat # N0440S), 5 µL $CoCl_2$, and incubated at 37 °C for 30 min. Then, nuclei were washed with PBST three times to wash away the residual reagents.

Design and Fabrication of the Microfluidic Device: The microfluidic chip shown in Figure S2a (Supporting Information) was designed using the computer-aided design software AutoCAD (2021, AutoDESK, USA). The established protocols[49] were employed for fabricating the polydimethylsiloxane (PDMS) microfluidic chip, with a channel depth of 50 µm. Molds for a microfluidic device were made using a photolithographic approach, consisting of centrifugal coating and modeling the SU-8. Silicon molds were employed for casting PDMS (Sylgard-184) to fabricate microfluidic devices.

AI-Assistant Droplets Barcoding: As illustrated in Figure 2c,d, an AI-assistant droplet barcoding system consisting of a home-made pressure pump was developed, a motorized linear slider with tube housing, an imaging unit, a main control panel, and a mechanical stand for microfluidic chip mounting. The homemade pressure pump was integrated within the AI-assistant droplet barcoding system and capable of supplying at least four independent pneumatic pressure sources with a range of 0–40 kPa and stability of <0.02 kPa. The imaging unit consists of an LED lighting source, a CMOS camera (MER2-U3, Daheng Imaging), and a 4X lens and was mounted underneath the droplet generation region of the microfluidic chip. Once the experiment was triggered, droplet generation within the microfluidic chip was real-time monitored and analyzed. The captured images for droplet generation were sent to the main control panel for processing and video clips were shown to the experiment operator via the man-machine interface. At the beginning and ending stages, unqualified droplet series were discarded into the waste tube. Once droplet quality was confirmed by the control panel, the motorized liner slider was switched, and therefore qualified droplets were collected to the keeping tube. After the experiment is completed, a summary report including a series of parameters, such as droplet counts, diameters, etc., will be shown in the man-machine interface, and/or a PDF version will be available for download.

Real-Time Image Processing: The dataset including more than 5000 droplet generation images was collected from the previous experiment and annotated manually. The computer vision model was then trained using YOLOv5 and verified using 1000+ additional experimental images. During the experimental course, real-time droplet generation was recorded by the imaging unit, and then droplet quality was analyzed and processed by a control panel using the trained model. Modeling training was carried out as follows. Step 1: Microdroplet morphology dataset creation. First, collect 5000+ various droplet generation images within the microfluidic channel using the droplet barcoding system. Second, manually label each collected image and identify the droplet subtype (for example, 1-droplet-0-bead, 1-droplet-1-bead, 1-droplet-2-beads, over-sized droplet, under-sized droplet, etc.) using labeling (an opensource tool). Step 2: Split the dataset and modify the configuration file. First, by running a self-written Python code, split the collected images to 3 subtypes, that is, testing set (10%), training set (80%), and verification set (10%), and save them into predefined catalogs. Second, create a new .yaml file under the data folder within the YOLOv5 directory, and then modify the created .yaml file to match the training according to YOLOv5 official instructions. Step 3: Model training. First, run train.py file within the YOLOv5 directory and save after adjusting relevant parameters to match the training. Second, rerun train.py file to trigger the training process and then a best.pt file would be created once completed. Step 4: Model verification. Verify the model using created best.pt file under: runs/train/exp/weights within the YOLOv5 directory. The verification results are available under YOLOV5/runs/detect and model accuracy will be summarized by comparing the verification results with original labeled images. Iteration of Step 2, 3, and 4 is necessary to continual improve the model until the model accuracy is stably >85% during the model verification.

Second-Strand cDNA Synthesis: The morphology of nuclei after in situ reactions was observed by optical microscope. To ensure a maximum of one nucleus per droplet, nuclei were counted and diluted to a concentration approximating one cell in every ≈10 droplets, based on the *Poisson* distribution described by Zilionis et al.[49] To prevent cell sedimentation in the microfluidic chip's input well during the droplet barcoding process, the nuclei were adjusted to match the medium's density by using a 50% Optiprep solution. Nuclei, 2X DNA extension reaction mix, and barcoded beads (M20 Genomics, Cat # R20123124) were encapsulated into droplets using the AI-assistant droplet barcoding system. Then, the emulsions were incubated at 37 °C for 1 h, 50 °C 30 min, 60 °C 30 min, and 75 °C 20 min. After the barcoding reaction, droplets were broken by mixing with PFO buffer. The aqueous phase was taken out and purified by Ampure XP beads (Beckmen, Cat #A63881). PCR primers (Primer1 and Primer2) were ordered from Sangon Biotech company (China) according to the sequence provided in the previous work.[9] PCR was performed to amplify the purified product using Primer1 and Primer2 primers. The amplified product was purified by Ampure XP beads and quantified by Qubit.

Library Preparation: The sequencing library was constructed according to the VAHTS Universal DNA Library Prep Kit (Vazyme, Cat #ND607-01) for Illumina V3. About 50 ng of DNA fragments were used to construct the sequencing library. The input-DNA and final library were quantified by Qubit2.0 (Life Technologies). The fragment sizes of input DNA and final library were measured with Qsep100 DNA Fragment Analyzer (BIOptic). The DNA fragments were purified and selected using AMPure XP beads. Sequencing was performed using the NovaSeq 6000 and S4 Reagent Kit with paired end reads of 150.

Data Analysis—Preprocessing of Automated snRandom-seq Data: First, primer sequences and extra bases generated by the dA-tailing step were trimmed in raw sequencing data. Then for each Read1, UMI (8 nts) and cell-specific barcode (30 nts) were extracted and merged sequenced barcodes that could be uniquely assigned to the same accepted barcode with a Hamming distance of 2 nts or less. Read2 was used to generate the gene expression matrix by the STARsolo module in STAR (2.7.10a) with reasonable parameters. To determine the number of nuclei in each sample, the scattergram of log10(genes) for each possible barcode was plotted and used the position of the minimum with the highest value of log10(genes) as the threshold: only barcodes with the number of genes above this threshold were used for downstream analysis.

Data Analysis—Clustering and Downstream Analysis: The gene expression matrix was generated after barcode filtering and removal of mitochondrial RNAs and ribosomal RNAs. The analysis and visualization of automated snRandom-seq data were conducted using the Seurat 3 toolkit[50] within RStudio (4.2.1), which encompassed a range of processes: preprocessing, integration, visualization, clustering, cell type identification, and differential expression testing. Genes detected in fewer than 3 nuclei were filtered out. The following thresholds were used for nuclei-level filtering: nCount_RNA < CountThresh & nFeature_RNA > 200. CountThresh = mean(nCount_RNA) + 2 * sd(nCount_RNA). For the integration of automated snRandom-seq datasets, counts were normalized and scaled in Seurat. The integration was executed using the Harmony package[51] in R. Integrations were performed across the FFPE/fresh comparison samples, tumor subtypes, and primary/recurrent comparison samples, respectively. Within each sample, 2000 anchors were identified, and the integration of automated snRandom-seq datasets was realized through the IntegrateData function, utilizing 20 dimensions.[52] To construct integrated datasets, the shared nearest neighbor (SNN) graph was created by conducting principal component analysis (PCA), followed by the application of FindNeighbors using 30 principal components. Clusters were subsequently delineated using the FindClusters function with a resolution of 1. The visualization of clusters used UMAP of the principal components, as implemented in Seurat. The cell type identification for each cluster was accomplished manually using a published set of marker genes. Marker genes were identified using the FindAllMarkers function within Seurat. The resultant marker genes matching the filter criteria (only.pos = TRUE, min.pct = 0.25, logfc.threshold = 0.25) were kept. lncRNA was defined as long noncoding RNA sequences identified from Gene Transfer Format (GTF) files downloaded from the GENCODE database. lncRNAs without associated gene names were classified as unannotated lncRNAs.

Data Analysis—Correlation Analysis: To compare gene expression levels with the scRNA-seq data from FFPE and fresh samples, standard normalization and scaling procedures were applied. The average normalized expression values were calculated using Seurat's "AverageExpression" function. The natural logarithm of the average expression with one added pseudo count was plotted and the coefficient of variation and *p*-value were calculated using the ggpubr package (0.4.0) in R.

Data Analysis—Functional Pseudotime Analysis: The differentiation trajectory of a set of glioma cells was performed with Monocle3 and Seurat packages in R.

Data Analysis—Analyses of Lipid Metabolic Pathways: The visualization and quantification of the lipid metabolic diversity of single cells in each cluster were performed with scMetabolism (v0.2.1)[53] package in R.

Data Analysis—Group Preference Analysis: Group preference of each cell state in primary-recurrent GBM samples was calculated by the Chi-Square test ($R_{O/E}$)[54] using chisq.test function in R.

Data availability: The automated snRandom-seq data generated in this study have been deposited in the Genome Sequence Archive database. Source data are provided with this paper.

Code availability: The code for the preprocessing of automated snRandom-seq data is available at https://github.com/wanglab2023/smRandom-seq.

Statistical Analysis: Statistical details for each experiment are provided in the figure legends. Pre-processing of the automated snRandom-seq data was described in Data Analysis section. The single nuclei isolation experiment and droplet barcoding experiment were repeated more than three times independently with similar results. The *p* value (*p*) for the Pearson's correlation coefficient (*R*) was computed from two-sided permutation test. The ggpubr package (0.4.0) in *R* was used for statistical analysis.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

The project was supported by the National Natural Science Foundation of China (No. 32200073, Y.W., 32250710678, Y.W. and No. 82200977, Z.X.), Leading Innovative and Entrepreneur Team Introduction Program of Zhejiang (No. 2021R01012, Y.W.), "Pioneer" R&D programs of Zhejiang Province (No. 2024C03005, Y.W.), and Key R&D Program of Zhejiang (No. 2024SSYS0022, Y.W.), Beijing Xisike Clinical Oncology Research Foundation (Y-zai2021/qn-0204 and Y-zai2021/zd-0207, Z.Q.). Thanks for the technical support from the core facilities of Zhejiang University and Liangzhu Laboratory. The authors thank Jingyao Chen and Chengcheng Zhang from the Core Facilities, Zhejiang University School of Medicine for their technical support.

Conflict of Interest

Y.W. is a co-founder, Q.L. is a co-founder and employee, and M.Z. is an employee of M20 Genomics. Y.S. is an employee of a startup developing medical devices for treating gliomas. The remaining authors declare no competing interests.

Author Contributions

Z.X., L.C., X.L., and Y.L. contributed equally to this work. Y.W., Z.Q., Q.L., and Y.S. conceived the study and managed the project's progress. Y.W. and Z.X. coordinated the experiments and analysis. Y.L., M.Z., H.Z., and Q.L. constructed the automated and AI-assistant platform. Z.X. and H.C. developed the automated snRandom-seq chemistry and performed experiments. Z.X., L.C., X.L., and T.Z. performed data analysis and visualization. L.C., X.L., Y.C., and Z.Q. collected samples and clinical information. Other authors assisted with the experiments and participated in critical discussions. Z.X. and Y.L. wrote the paper. All authors have revised and approved the final manuscript.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Keywords

automated, FFPE, Gliomas, lncRNA, snRNA-seq

Received: December 29, 2023
Revised: June 20, 2024
Published online:

---

[1] M. L. Goodenberger, R. B. Jenkins, *Cancer Genet.* **2012**, *205*, 613.

[2] Q. T. Ostrom, L. Bauchet, F. G. Davis, I. Deltour, J. L. Fisher, C. E. Langer, M. Pekmezci, J. A. Schwartzbaum, M. C. Turner, K. M. Walsh, M. R. Wrensch, J. S. Barnholtz-Sloan, *Neuro Oncol.* **2014**, *16*, 896.

[3] L. R. Schaff, I. K. Mellinghoff, *JAMA, J. Am. Med. Assoc.* **2023**, *329*, 574.

[4] R. A. Burrell, N. McGranahan, J. Bartek, C. Swanton, *Nature* **2013**, *501*, 338.

[5] K. Wang, T. Kumar, J. Wang, D. C. Minussi, E. Sei, J. Li, T. M. Tran, A. Thennavan, M. Hu, A. K. Casasent, Z. Xiao, S. Bai, L. Yang, L. M. King, V. Shah, P. Kristel, C. L. van der Borden, J. R. Marks, Y. Zhao, A. J. Zurita, A. Aparicio, B. Chapin, J. Ye, J. Zhang, D. L. Gibbons, E. Sawyer, A. M. Thompson, A. Futreal, E. S. Hwang, J. Wesseling, et al., *Cell* **2023**, *186*, 3968.

[6] D. L. Evers, J. He, Y. H. Kim, J. T. Mason, T. J. O'Leary, *J. Mol. Diagn.* **2011**, *13*, 687.

[7] F. V. Andres, H. Kate, W. Taopeng, W. Kellie, M. B. Lisa, P. Jose, P. Jasmine, S. Alex, G. M. Luciano, *bioRxiv* **2022**, bioRxiv:08.23.505054.

[8] C. Hattie, M. Alexandre, M. Cristin, D. Eugene, W. Nicholas Van, M. M. Emma, W. Julia, S. Avrum, C. Fei, M. Sarah, R.-R. Orit, R. Aviv, *bioRxiv* **2022**, bioRxiv:08.25.505257.

[9] Z. Xu, T. Zhang, H. Chen, Y. Zhu, Y. Lv, S. Zhang, J. Chen, H. Chen, L. Yang, W. Jiang, S. Ni, F. Lu, Z. Wang, H. Yang, L. Dong, F. Chen, H. Zhang, Y. Chen, J. Liu, D. Zhang, L. Fan, G. Guo, Y. Wang, *Nat. Commun.* **2023**, *14*, 2734.

[10] G. X. Y. Zheng, J. M. Terry, P. Belgrader, P. Ryvkin, Z. W. Bent, R. Wilson, S. B. Ziraldo, T. D. Wheeler, G. P. McDermott, J. Zhu, M. T. Gregory, J. Shuga, L. Montesclaros, J. G. Underwood, D. A. Masquelier, S. Y. Nishimura, M. Schnall-Levin, P. W. Wyatt, C. M. Hindson, R. Bharadwaj, A. Wong, K. D. Ness, L. W. Beppu, H. J. Deeg, C. McFarland, K. R. Loeb, W. J. Valente, N. G. Ericson, E. A. Stevens, J. P. Radich, et al., *Nat. Commun.* **2017**, *8*, 14049.

[11] F. Salmen, J. De Jonghe, T. S. Kaminski, A. Alemany, G. E. Parada, J. Verity-Legg, A. Yanagida, T. N. Kohler, N. Battich, F. van den Brekel, A. L. Ellermann, A. M. Arias, J. Nichols, M. Hemberg, F. Hollfelder, A. van Oudenaarden, *Nat. Biotechnol.* **2022**, *40*, 1780.

[12] A. Bhaduri, E. Di Lullo, D. Jung, S. Müller, E. E. Crouch, C. S. Espinosa, T. Ozawa, B. Alvarado, J. Spatazza, C. R. Cadwell, G. Wilkins, D. Velmeshev, S. J. Liu, M. Malatesta, M. G. Andrews, M. A. Mostajo-Radji, E. J. Huang, T. J. Nowakowski, D. A. Lim, A. Diaz, D. R. Raleigh, A. R. Kriegstein, *Cell Stem Cell* **2020**, *26*, 48.

[13] C. Neftel, J. Laffy, M. G. Filbin, T. Hara, M. E. Shore, G. J. Rahme, A. R. Richman, D. Silverbush, M. L. Shaw, C. M. Hebert, J. Dewitt, S. Gritsch, E. M. Perez, L. N. Gonzalez Castro, X. Lan, N. Druck, C. Rodman, D. Dionne, A. Kaplan, M. S. Bertalan, J. Small, K. Pelton, S. Becker, D. Bonal, Q. D. Nguyen, R. L. Servis, J. M. Fung, R. Mylvaganam, L. Mayr, J. Gojo, et al., *Cell* **2019**, *178*, 835.

[14] C. Guetta-Terrier, D. Karambizi, B. Akosman, J. P. Zepecki, J. S. Chen, S. Kamle, J. E. Fajardo, A. Fiser, R. Singh, S. A. Toms, C. G. Lee, J. A. Elias, N. Tapinos, *Cancer Res.* **2023**, *83*, 1984.

[15] A. Mamivand, S. Bayat, A. Maghrouni, S. Shabani, A. Khoshnevisan, H. Saffar, M. Tabrizi, *Clin. Epigenetics* **2022**, *14*, 35.

[16] S. M. Robert, H. Sontheimer, *Cell. Mol. Life Sci.* **2014**, *71*, 1839.

[17] X. Fei, Y. N. Dou, K. Sun, J. Wei, Q. Guo, L. Wang, X. Wu, W. Lv, X. Jiang, Z. Fei, *Exp. Mol. Med.* **2023**, *55*, 1203.

[18] C. Zhu, J. M. Kros, C. Cheng, D. Mustafa, *Neuro Oncol.* **2017**, *19*, 1435.

[19] S. Jain, J. W. Rick, R. S. Joshi, A. Beniwal, J. Spatz, S. Gill, A. C. Chang, N. Choudhary, A. T. Nguyen, S. Sudhir, E. J. Chalif, J. S. Chen, A. Chandra, A. F. Haddad, H. Wadhwa, S. S. Shah, S. Choi, J. L. Hayes, L. Wang, G. Yagnik, J. F. Costello, A. Diaz, D. H. Heiland, M. K. Aghi, *J. Clin. Invest* **2023**, *133*, 147087.

[20] Y. Sugita, J. J. Kepes, M. Shigemori, S. Kuramoto, G. Reifenberger, J. C. Kiwit, W. Wechsler, *Clin. Neuropathol.* **1990**, *9*, 271.

[21] S. M. Mousavi, M. Derakhshan, F. Baharloii, F. Dashti, S. M. A. Mirazimi, M. Mahjoubin-Tehran, S. Hosseindoost, P. Goleij, N. Rahimian, M. R. Hamblin, H. Mirzaei, *Mol. Ther. Oncolytics* **2022**, *24*, 262.

[22] A. Goenka, X. Song, D. Tiek, R. P. Iglesia, M. Lu, C. Zeng, C. Horbinski, W. Zhang, B. Hu, S. Y. Cheng, *Neuro Oncol.* **2023**, *25*, 1592.

[23] T. Peng, D. L. Chen, S. L. Chen, *Bioengineered* **2022**, *13*, 9172.

[24] X. Liu, Q. Zhu, Y. Guo, Z. Xiao, L. Hu, Q. Xu, *Biomed. Pharmacother.* **2019**, *117*, 109069.

[25] J. Liang, C. Liu, D. Xu, K. Xie, A. Li, *J. Transl. Med.* **2022**, *20*, 80.

[26] S. M. Robert, S. C. Buckingham, S. L. Campbell, S. Robel, K. T. Holt, T. Ogunrinu-Babarinde, P. P. Warren, D. M. White, M. A. Reid, J. M. Eschbacher, M. E. Berens, A. C. Lahti, L. B. Nabors, H. Sontheimer, *Sci. Transl. Med.* **2015**, *7*, 289ra86.

[27] A. Monnier, R. Boniface, R. Bouvet, A. Etcheverry, M. Aubry, T. Avril, V. Quillien, E. Chevet, J. Mosser, *BMC Cancer* **2018**, *18*, 1213.

[28] T. Shingu, A. L. Ho, L. Yuan, X. Zhou, C. Dai, S. Zheng, Q. Wang, Y. Zhong, Q. Chang, J. W. Horner, B. D. Liebelt, Y. Yao, B. Hu, Y. Chen, G. N. Fuller, R. G. W. Verhaak, A. B. Heimberger, J. Hu, *Nat. Genet.* **2017**, *49*, 75.

[29] Y. Li, J. Xu, J. Zhang, J. Zhang, J. Zhang, X. Lu, *Cancer Cell Int.* **2019**, *19*, 294.

[30] W. Z. Hou, X. L. Chen, L. S. Qin, Z. J. Xu, G. M. Liao, D. Chen, L. J. Hu, Z. M. Mao, J.-S. Huang, Q. Yuan, C. Liao, Q. M. Yao, *Riv. Eur. Sci. Med. Farmacol.* **2020**, *24*, 5549.

[31] H. Wu, C. Guo, C. Wang, J. Xu, S. Zheng, J. Duan, Y. Li, H. Bai, Q. Xu, F. Ning, F. Wang, Q. Yang, *Cancer Sci.* **2023**, *114*, 2609.

[32] C. Lu-Emerson, M. Snuderl, N. D. Kirkpatrick, J. Goveia, C. Davidson, Y. Huang, L. Riedemann, J. Taylor, P. Ivy, D. G. Duda, M. Ancukiewicz, S. R. Plotkin, A. S. Chi, E. R. Gerstner, A. F. Eichler, J. Dietrich, A. O. Stemmer-Rachamimov, T. T. Batchelor, R. K. Jain, *Neuro Oncol.* **2013**, *15*, 1079.

[33] A. Makarevic, C. Rapp, S. Dettling, D. Reuss, C. Jungk, A. Abdollahi, A. von Deimling, A. Unterberg, C. Herold-Mende, R. Warta, *Int. J. Mol. Sci.* **2020**, *21*, 7801.

[34] L. Wu, W. Wu, J. Zhang, Z. Zhao, L. Li, M. Zhu, M. Wu, F. Wu, F. Zhou, Y. Du, R.-C. Chai, W. Zhang, X. Qiu, Q. Liu, Z. Wang, J. Li, K. Li, A. Chen, Y. Jiang, X. Xiao, H. Zou, R. Srivastava, T. Zhang, Y. Cai, Y. Liang, B. Huang, R. Zhang, F. Lin, L. Hu, X. Wang, et al., *Cancer Discovery* **2022**, *12*, 2820.

[35] M. G. Filbin, I. Tirosh, V. Hovestadt, M. L. Shaw, L. E. Escalante, N. D. Mathewson, C. Neftel, N. Frank, K. Pelton, C. M. Hebert, C. Haberler, K. Yizhak, J. Gojo, K. Egervari, C. Mount, P. van Galen, D. M. Bonal, Q. D. Nguyen, A. Beck, C. Sinai, T. Czech, C. Dorfer, L. Goumnerova, C. Lavarino, A. M. Carcaboso, J. Mora, R. Mylvaganam, C. C. Luo, A. Peyrl, M. Popović, et al., *Science* **2018**, *360*, 331.

[36] Z. Li, R. Qian, J. Zhang, X. Shi, *Biosci. Rep.* **2019**, *39*, BSR20180879.

[37] Y. Li, W. Li, X. Zeng, X. Tang, S. Zhang, F. Zhong, X. Peng, Y. Zhong, T. J. Rosol, X. Deng, Z. Liu, S. Peng, X. Peng, *Oncogene* **2019**, *38*, 7234.

[38] K. Abdul Rashid, K. Ibrahim, J. H. D. Wong, N. Mohd Ramli, *Metabolites* **2022**, *12*, 1280.

[39] A. P. Patel, I. Tirosh, J. J. Trombetta, A. K. Shalek, S. M. Gillespie, H. Wakimoto, D. P. Cahill, B. V. Nahed, W. T. Curry, R. L. Martuza, D. N. Louis, O. Rozenblatt-Rosen, M. L. Suvà, A. Regev, B. E. Bernstein, *Science* **2014**, *344*, 1396.

[40] C. P. Couturier, S. Ayyadhury, P. U. Le, J. Nadaf, J. Monlong, G. Riva, R. Allache, S. Baig, X. Yan, M. Bourgey, C. Lee, Y. C. D. Wang, V. Wee Yong, M.-C. Guiot, H. Najafabadi, B. Misic, J. Antel, G. Bourque, J. Ragoussis, K. Petrecca, *Nat. Commun.* **2020**, *11*, 3406.

[41] N. Ochocka, P. Segit, K. A. Walentynowicz, K. Wojnicki, S. Cyranowski, J. Swatler, J. Mieczkowski, B. Kaminska, *Nat. Commun.* **2021**, *12*, 1151.

[42] N. D. Mathewson, O. Ashenberg, I. Tirosh, S. Gritsch, E. M. Perez, S. Marx, L. Jerby-Arnon, R. Chanoch-Myers, T. Hara, A. R. Richman, Y. Ito, J. Pyrdol, M. Friedrich, K. Schumann, M. J. Poitras, P. C. Gokhale, L. N. Gonzalez Castro, M. E. Shore, C. M. Hebert, B. Shaw, H. L. Cahill, M. Drummond, W. Zhang, O. Olawoyin, H. Wakimoto, O. Rozenblatt-Rosen, P. K. Brastianos, X. S. Liu, P. S. Jones, D. P. Cahill, et al., *Cell* **2021**, *184*, 1281.

[43] T. A. Steiert, G. Parra, M. Gut, N. Arnold, J.-R. Trotta, R. Tonda, A. Moussy, Z. Gerber, P. M. Abuja, K. Zatloukal, C. Röcken, T. Folseraas, M. M. Grimsrud, A. Vogel, B. Goeppert, S. Roessler, S. Hinz, C. Schafmayer, P. Rosenstiel, J.-F. Deleuze, I. G. Gut, A. Franke, M. Forster, *Nucleic Acids Res.* **2023**, *51*, 7143.

[44] A. Alavi, M. Ruffalo, A. Parvangada, Z. Huang, Z. Bar-Joseph, *Nat. Commun.* **2018**, *9*, 4768.

[45] M. Guo, M. P. Morley, C. Jiang, Y. Wu, G. Li, Y. Du, S. Zhao, A. Wagner, A. C. Cakar, M. Kouril, K. Jin, N. Gaddis, J. A. Kitzmiller, K. Stewart, M. C. Basil, S. M. Lin, Y. Ying, A. Babu, K. A. Wikenheiser-Brokamp, K. S. Mun, A. P. Naren, G. Clair, J. N. Adkins, G. S. Pryhuber, R. S. Misra, B. J. Aronow, T. L. Tickle, N. Salomonis, X. Sun, E. E. Morrisey, et al., *Nat. Commun.* **2023**, *14*, 4566.

[46] R. Qi, Q. Zou, *Research* **2023**, *6*, 0050.

[47] Y. Chen, Q. Zhan, J. Zhang, W. Wang, B. Luan Khoo, Z. Liu, S. Wei, J. Niu, J. Xu, C. C. Yu, X. Hu, Y. Liu, J. Han, S. Liu, L. Liu, *Anal. Chim. Acta* **2022**, *1221*, 340151.

[48] D. S. Veliz, K. L. Lin, C. Sahlgren, *Smart Med.* **2023**, *2*, 20220030.

[49] R. Zilionis, J. Nainys, A. Veres, V. Savova, D. Zemmour, A. M. Klein, L. Mazutis, *Nat. Protoc.* **2017**, *12*, 44.

[50] C. Hafemeister, R. Satija, *Genome Biol.* **2019**, *20*, 296.

[51] I. Korsunsky, N. Millard, J. Fan, K. Slowikowski, F. Zhang, K. Wei, Y. Baglaenko, M. Brenner, P. R. Loh, S. Raychaudhuri, *Nat. Methods* **2019**, *16*, 1289.

[52] T. Stuart, A. Butler, P. Hoffman, C. Hafemeister, E. Papalexi, W. M. Mauck, 3rd, Y. Hao, M. Stoeckius, P. Smibert, R. Satija, *Cell* **2019**, *177*, 1888.

[53] Y. Wu, S. Yang, J. Ma, Z. Chen, G. Song, D. Rao, Y. Cheng, S. Huang, Y. Liu, S. Jiang, J. Liu, X. Huang, X. Wang, S. Qiu, J. Xu, R. Xi, F. Bai, J. Zhou, J. Fan, X. Zhang, Q. Gao, *Cancer Discovery* **2022**, *12*, 134.

[54] X. Guo, Y. Zhang, L. Zheng, C. Zheng, J. Song, Q. Zhang, B. Kang, Z. Liu, L. Jin, R. Xing, R. Gao, L. Zhang, M. Dong, X. Hu, X. Ren, D. Kirchhoff, H. G. Roider, T. Yan, Z. Zhang, *Nat. Med.* **2018**, *24*, 978.