

scONE-seq: A one-tube single-cell multi-omics method enables simultaneous dissection of molecular phenotype and genotype heterogeneity from frozen tumors

Angela Wu (✉ angelawu@ust.hk)

Hong Kong University of Science and Technology <https://orcid.org/0000-0002-3531-4830>

Lei Yu

Hong Kong University of Science and Technology

Xinlei Wang

Hong Kong University of Science and Technology

Quanhua Mu

Hong Kong University of Science and Technology

Sindy Tam

Hong Kong University of Science and Technology

Danson Loi

Hong Kong University of Science and Technology

Aden Chan

The Chinese University of Hong Kong

Wai Sang Poon

The Chinese University of Hong Kong

Ho-Keung Ng

The Chinese University of Hong Kong

Danny Chan

The Chinese University of Hong Kong

Jiguang Wang

Hong Kong University of Science and Technology

Article

Keywords: scONE-seq, single-cell-whole-genome-RNA-sequencing, heterogeneity

Posted Date: October 5th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-956200/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Genomic and transcriptomic heterogeneity both play important roles in normal cellular function as well as in disease development. To be able to characterize these different forms of cellular heterogeneity in diverse sample types, we developed scONE-seq, which enables simultaneous transcriptome and genome profiling in a one-tube reaction. Previous single-cell-whole-genome-RNA-sequencing (scWGS-RNA-seq) methods require physical separation of DNA and RNA, often by physical separation of the nucleus from the cytoplasm. These methods are labor-intensive and technically demanding, time-consuming, or require special devices, and they are not applicable to frozen samples that cannot generate intact single-cell suspensions. scONE-seq is a one-tube reaction, thus is highly scalable and is the first scWGS-RNA-seq method compatible with frozen biobanked tissue. We benchmarked scONE-seq against existing methods using cell lines and lymphocytes from a healthy donor, and we applied it to a 2-year-frozen astrocytoma sample profiling over 1,200 nuclei, subsequently identifying a unique transcriptionally normal-like tumor clone. scONE-seq makes it possible to perform large-scale single-cell multi-omics interrogation with ease on the vast quantities of biobanked tissue, which could transform the scale of future multi-omics single-cell cancer profiling studies.

Introduction

Single-cell genomics has become a mainstay technology used to dissect multicellular organisms and tissues that are composed of cells with diverse functions¹⁻⁴. The power of this approach has been demonstrated in several cell atlas studies: novel cell types have been discovered that further led to the elucidation of new mechanisms; complex cellular interactions and transitions associated with disease initiation or progression have been revealed; cross-species analyses have shed light on evolutionary processes⁵⁻⁸. The use of single-cell technology in studying cancer is especially important. Regulatory mechanisms underlying drug resistance or immune evasion are elusive and complex, and tumor cell heterogeneity is a major contributing factor to this complexity, making it particularly challenging to dissect these mechanisms with bulk techniques⁹⁻¹¹. Single-cell technologies have greatly enhanced our understanding of tumor heterogeneity and accelerated mechanistic discovery. At the phenotype level, single-cell RNA-seq (scRNA-seq) has been used to uncover drug-resistant melanoma subpopulations and to characterize cancer stem cell subpopulations in glioblastoma^{7,12-15}. scRNA-seq has also enabled a more comprehensive phenotypic understanding of the tumor microenvironment (TME) in many cancers including glioma and colorectal cancer¹⁶⁻²⁰. At the genotype level, genomic instability contributes to cancer initiation, progression, relapse, and metastasis⁹. With single-cell whole genome sequencing (scWGS), the clonal structure of the tumor can be resolved, and evolutionary analysis based on copy number aberrations (CNAs) can reveal tumor progression^{21,22}.

Evidently, both the genomic and transcriptomic heterogeneity of tumors contributes to the disease, and understanding the importance of both in cancer studies is crucial. Several single-cell methods that interrogate DNA and RNA simultaneously in the same cell (scWGS-RNA-seq) have been developed²³⁻²⁸. However, these first-generation scWGS-RNA-seq methods have not been widely applied since their invention, largely because they require physical separation of DNA and RNA, often by physical separation of the nucleus from the cytoplasm. These separation techniques are labor-intensive and technically demanding, time-consuming, requiring well-trained experimental techniques, or require special microfluidic devices^{29,30}. Furthermore, they are not applicable to frozen samples, where it is impossible to obtain intact single-cell suspensions. As such, existing scWGS-RNA-seq methods cannot be applied to the vast majority of primary biobanked tumor samples, which represents most of the readily available clinical samples. All in all, these limitations make first-generation scWGS-RNA-seq methods not easily accessible.

We introduce a novel single-cell DNA and RNA co-amplification method, scONE-seq, which enables co-profiling of the transcriptome and genome from the same single cell or nucleus in a one-tube reaction. We devised a barcoding strategy that introduces a 6-base long DNA-specific and RNA-specific barcodes to each type of nucleic acid during the single-cell DNA/RNA amplification process, while also incorporating UMIs³¹⁻³³. Thus, DNA and RNA reads can be distinguished by their barcode information after sequencing. scONE-seq has several advantages: it has a simplified library construction workflow; it is compatible with standard biology workflows such as fluorescence-activated cell sorting (FACS); being a one-pot reaction, its throughput can be easily scaled up using liquid-handling robots; most importantly, scONE-seq does not require physically separating the DNA and RNA, and is therefore applicable to a variety of sample types including single nuclei. This means that frozen clinical samples and tissue types like liver, bone, and brain, that are difficult to dissociate into single-cell suspensions, which are intractable with other scWGS-RNA-seq methods, can be profiled using scONE-seq.

We demonstrate that the technical performance of scONE-seq is comparable to existing single-cell methods in a benchmarking experiment using various sample types, including several cell lines and lymphocytes from the blood of a healthy donor. We then show how scONE-seq can be used to discover novel disease-related information by applying it to a frozen IDH1-mutant astrocytoma sample. Using scONE-seq data, we were able to simultaneously characterize both the molecular phenotype and genotype heterogeneity of this patient sample, profiling the cell types in the tumor as well as the clonal events. We were able to fully dissect the genome level and transcriptomic level heterogeneity of this patient's disease, and in doing so we identified a unique normal-like tumor clone by integration of clonal and transcriptomic information. Further analysis of RNA and DNA profiles of this clone suggests that this subpopulation may play a role in the maintenance of the tumor, as well as in regulating the tumor's interactions with normal cells in its microenvironment. The abnormal genomic profile of this small subpopulation along with its normal-like transcriptomic phenotype hints at its origins during the tumor initiation process and demonstrates the power of scONE-seq to distinguish genomic variations of near-normal cells from true normal cells. Using this frozen primary glioma sample as an example, we establish a framework for integrative single-cell multi-omic WGS/RNA-seq analysis on biobanked samples to generate new insights on tumor heterogeneity, enabled by the versatility and scalability of the scONE-seq approach.

Results

Molecular barcoding strategy enables accurate and sensitive co-profiling of DNA and RNA from a single cell in a one-tube reaction

To achieve single-cell genome and transcriptome co-profiling, we devised a workflow to amplify RNA and DNA simultaneously (Fig. 1a). Briefly, to perform scONE-seq, after the sample dissociation, cells or nuclei are sorted into PCR plates containing lysis buffer with a flow cytometer. Plates of sorted single cells can then be processed immediately, or store at -80 for months before processing. To start the single-cell amplification, we first use Tn5 with a custom adaptor to fragment and label the genome or any other DNA within the cells^{34–36}. In this step, the amplification adaptor, which includes a 6-nucleotide “DNA barcode” and 6-nucleotide UMI, is added to fragmented DNA (fDNA). Subsequently, we use reverse transcription (RT) to generate cDNA, where the RT primer is comprised of: a priming sequence adapted from that of the MATQ-seq protocol³⁷; a 6-nucleotide “RNA barcode”; and a 5-nucleotide UMI. The RT priming sequence is a modified random oligo and primes to the internal regions of RNA transcripts, thereby enabling detection of full-length transcripts, including non-polyadenylated (non-polyA) RNAs. cDNA 3’ adaptor is then added through subsequent poly-C tailing and degenerate PCR^{38,39}. Once DNA-specific and RNA-specific barcodes have been added, fDNA and cDNA are amplified simultaneously, and the sequencing library is constructed with the pre-amplified products (Methods).

First, to characterize the transcriptome generated by scONE-seq, we benchmarked it against Smart-seq2 (SS2)^{40,41} using a variety of test samples: extracted RNA-free *E. coli* genomes (mock DNA), extracted DNA-free human total RNA (mock RNA), as well as a mixture of the two (i.e. *E. coli* DNA mixed with human total RNA); and cultured HCT116 single cells. We evaluated the sensitivity by assessing the number of genes detected in each of the benchmark mock and HCT116 samples and found that scONE-seq detected more genes per cell than SS2 (Fig. 1b and Supplementary Fig. 1a, $P < 2 \times 10^{-16}$, t-test). This is likely due to scONE-seq being able to capture total RNA^{37,42} while SS2 only targets polyA RNA, and therefore capturing a more diverse set of molecules at any given sequencing depth (Supplementary Fig. 1d). Also, scONE-seq enables full-length transcript profiling and achieves gene body coverage uniformity comparable to SS2 (Fig. 1c). We then used sample-to-sample correlation analysis as well as the detection of ERCC Spike-In to estimate the accuracy of scONE-seq in comparison to SS2. The sample-to-sample correlation analysis shows the two methods are comparable (Fig. 1d, Supplementary Fig. 1c). Using ERCC Spike-In as a measure of accuracy, SS2 performs slightly better with a correlation 0.91 (R value) against the expected ERCC concentrations, while scONE-seq is 0.86; both are sufficiently high for quantitative measurement of transcript abundance from single cells (Supplementary Fig. 1b).

Next, we sought to validate the whole genome sequencing (WGS) capability of scONE-seq. Lorenz curves⁴³ compare the coverage uniformity for each method, showing a good performance by scONE-seq (Fig. 1e). Then, we used bulk HCT116 WGS data in comparison to scWGS data generated by scONE-seq to confirm that CNAs captured by scONE-seq are consistent with those defined by bulk (5×10^6 cells) and pseudo bulk (86 cells) (Fig. 1f, g, h). In addition, we performed UMI deduplication on the scONE-seq DNA dataset, since our method adds UMI to the DNA fragments during the tagmentation step and found that deduplication successfully reduces the bias introduced during single-cell DNA amplification (Methods). To confirm this effect, we calculated the coefficient of variation (CV) for the raw genomic bin counts (genome bin size of 500kb), and for the UMI deduplicated data, and indeed the CV was significantly reduced upon UMI deduplication (Supplementary Fig. 1f, $P = 1.2 \times 10^{-7}$, t-test). These analyses show that scONE-seq captures true CNAs from single cells at 500-kb resolution.

Summarily, the analysis of scRNA-seq and scWGS data generated using benchmark samples shows that scONE-seq can profile genome and transcriptome data from the same single cell without compromising data quality as compared to existing standard methods.

scONE-seq data correctly assigns cell types from primary donor samples

After thoroughly assessing the technical performance of scONE-seq, we next applied it to known biologically heterogeneous samples to evaluate whether it can accurately identify cellular subtypes within a mixed population. To do so, we performed scONE-seq on four different cell lines, as well as on a primary peripheral blood mononuclear cell (PBMC) sample from a healthy donor.

First, we analyzed the cell-line dataset containing 86 HCT116 cells, 143 NPC43 cells, 37 HUVEC cells, and 17 H9 cells to check for accurate cell-type assignment. With unsupervised graph-based clustering of the RNA expression data, cells from the same cell lines successfully clustered together (Fig. 2a). We also checked the gene markers for each cell line (Fig. 2b) and notably, several well-studied gene markers for these cell lines are found in the scONE-seq dataset (Fig. 2b, Supplementary Fig. 2a).

Next, we used lymphocytes from PBMC to test scONE-seq cell-types clustering accuracy in primary samples. We prepared sequencing libraries with scONE-seq and SS2 from the same PBMC sample for comparison. After quality control filtering to remove low-quality cells, we collected 200 cells for scONE-seq and 194 cells for SS2. With unsupervised graph-based clustering, we found no difference in the cell-type composition between the two methods (Fig. 2c,e, $P = 0.5109$, Chi-squared test). After clustering, we annotated the cell types using known lymphocyte markers^{44–46} (Fig. 2d,f, and Supplementary Fig. 2b): B cells distinguished by *CD19* and *MS4A1* (CD20); T cells characterized by *CD3E*; Less differentiated T cells (Naive and memory T cells) identified by *SELL* (CD62L), *CCR7*, and *LEF1*; CD4+ T cells characterized by *CD4*; CD8+ T cells characterized by *CD8A* and *CD8B*; and cytotoxic T cells distinguished by *PRF1* and *NKG7*. Notably, cytotoxic T cells comprise both Gamma delta T cells ($\gamma\delta$ T cell; expressing *TRDC*, *TRGC1*, and *TRGC2*, Supplementary Fig. 2c) as well as effector memory T cells (TEM; lacking CCR7 expression and positively expressing *IL2RB*) (Supplementary Fig. 2b). We surmised that these are TEM rather than effector T cells since the sample is from a healthy donor. Beyond that, in the scONE-seq dataset, we also captured some regulatory T cells (Treg cell; *FOXP3+*, *CCR4+*) (Supplementary Fig. 2d) and detected several non-polyA genes including *PZP* and *SESN3* whose expressions in T-cells have previously been described⁴⁷ (Supplementary Fig. 2e); these features were not found in the SS2 dataset.

These results collectively demonstrate that scONE-seq RNA data can accurately capture the biological variation within a heterogeneous sample.

scONE-seq data identifies distinct clones in different samples

The analysis above shows the feasibility of scONE-seq for the cell-type assignment using RNA data. Next, we evaluated the performance of clone identification with scONE-seq WGS data. Here, we utilized scONE-seq WGS data that was obtained simultaneously from the cell lines used in the previous cell-type assignments analysis, and delineated the CNAs clonal structure of all four cell lines, followed by hierarchical clustering with their integer copy number profiles (Fig. 2g). From this analysis, we see that HCT116 maintains a relatively homogeneous clonal composition, whereas NPC43, a primary patient-derived cell line that shows strong genome instability, is comprised of 3 main clones (Fig. 2g). Furthermore, the CNAs structure of these 3 clones differ substantially compared to when the cell line was first established⁴⁸ (Fig. 2h, Supplementary Fig. 3a, b, c, d), especially in chromosomes 1, 3, 4, 6, 7, and 11. Correspondingly, distinctions between clones are mainly found in chromosomes 1, 3, 7, and 11 (Supplementary Fig. 3a, b, c, d). Based on this observation, the change in chromosome copy numbers during cell culturing of primary cell lines could be a common phenomenon in cell lines with abundant CNAs and unstable genomes. Studies have shown extensive genetic variation across different cell culture lines, and that single cells from some cell lines can give rise to populations with multiple clones due to genome instability^{21,49}. Additionally, with the matched transcriptomes of every single cell and their corresponding copy number states, we mapped the clonal information to the transcriptome UMAP for NPC43 and found that the CNAs in NPC43 did not impact the transcriptome state dramatically (Supplementary Fig. 3e). This demonstrates that scONE-seq can identify both phenotype and genotype states for each individual cell.

Dissecting the clonal structure and cell-type subpopulations of an IDH1-mutant astrocytoma

Gliomas, especially high grade ones, are some of the most aggressive malignant tumors originating in the brain^{50,51}. When studying gliomas or other brain tissues using single-cell technology, it is challenging to obtain intact dissociated whole single cells, especially cells with complex morphology, and could lead to biases in cell-type sampling⁵². As such, for brain single-cell profiling, single nucleus isolation is more widely used. To profile both the genotypic and phenotypic heterogeneity in a biobanked sample, we apply scONE-seq on single nuclei isolated from a snap-frozen astrocytoma specimen that has been stored since the surgery for two years prior to analysis: a second recurrent (2R) astrocytoma sample with *IDH1* (R132H), *TP53* (P278S), and *ATRX* (R781*) mutations (Supplementary Fig. 4a). The primary (P) and first recurrent (1R) samples were limited in quantity, and subject to whole exome sequencing (WES) and RNA sequencing in bulk⁵³ (Fig. 3a). In total, we used scONE-seq to profile over 1,200 nuclei, including 1,210 scRNA datasets, 1,089 scWGS datasets, generating 908 passed-QC paired DNA and RNA datasets.

First, we delineated the clonal structure of this 2R astrocytoma sample. Using dimension reduction on scWGS data, we clustered cells into four distinct genomic states (Fig. 3b), consisting of one cluster of normal cells and three tumor clones (Supplementary Fig. 5a). Whole genome duplication was also found in this tumor and validated by measuring each cell's DAPI intensity using flow cytometry (Supplementary Fig. 5b). We were also able to identify aneuploidy and loss of heterozygosity in multiple loci using the B-allele frequency estimation (Supplementary Fig. 5c, d), and the integer copy number of cells was subsequently calculated (Fig. 3c). Based on the genomic profile of each clone and the WES data from the primary and recurrent tumors, the phylogenetic tree of the patient was constructed. The 2R clone 1 was found to be closer to the root (normal cell), with fewer loss of heterozygosity (LOH) events, and it has very similar genome alterations as the primary tumor WES data (Fig. 3d, e; Supplementary Fig. 5e). The 2R clone 2 and clone 3 harbor many of the same deletion regions as the 1R tumor, resulting in LOH (Fig. 3e; Supplementary Fig. 5e). We surveyed the known driver genes and found *BRAF*, *MET*, and *MYC* to be amplified in all the 2R clones⁵³⁻⁵⁵ (Fig. 3b). Several key deletion events were found to only occur in the 2R clone 2 and clone 3, including deletion of *CDKN2A*, and *PTEN*; in particular, the homozygous deletion of *CDKN2A* is a known prognostic factor for IDH mutant astrocytomas, and was recently incorporated into the World Health Organization central nervous system tumors classification scheme as a sufficient criterion for classifying cases as Grade 4, even when classical histology criteria are not fulfilled⁵⁶ (Supplementary Fig. 5f).

Next, we analyzed the RNA data from this dataset. First, we performed unsupervised graph-based clustering on scONE-seq RNA data, obtaining multiple cell clusters that were then annotated based on their RNA markers. We found this tumor contains tumor-associated macrophages (TAM), neurons, astrocytes, oligodendrocytes, and tumor cells defined by their canonical cell type gene signatures (Fig. 4a, b). The complex tumor microenvironment (TME) indicates the highly infiltrated phenotype of this recurrence tumor. The tumor cells – those that have transcriptomes classified as tumor-like – display high *EGFR* expression, a well-known feature of high grade gliomas. These tumor cells can further be subset into 4 cellular states based on meta-module scores described by Neftel et al.¹³ (Fig. 4a): oligodendrocyte progenitor cell-like (OPC-like), neural progenitor cell-like (NPC-like), mesenchymal-like (MES-like), and astrocyte-like (AC-like) cellular states.

In addition to the phylogenetic tree obtained from DNA data, which dissects the clonality, we are also able to use paired RNA data to superimpose the cell-type information onto the clonal information to identify clonal subpopulations with unique functional, phenotypic features. To do so, we mapped the clonal information to the RNA UMAP to visualize the clonal distribution among different cell types (Fig. 4c). The 2R clone 3 is the major clone of this tumor and is differentiated into all 4 tumor phenotypes: OPC-like, NPC-like, MES-like, and AC-like cellular states. The 2R clone 2 consists predominantly of AC-like cells. The 2R clone 1 is the most interesting: using RNA data alone, all cells from this clone were clustered with normal astrocytes, indicating transcriptome similarity between 2R clone 1 and normal astrocytes that is indistinguishable using scRNA-seq data alone; but upon superimposing matched genotype and phenotype information, this unique population of astrocyte-like tumor cells with clearly abnormal genotype as compared to the true normal cells are revealed (Fig. 4c). This highlights the importance of directly measuring copy number profiles as a standard to identify cancer cells and is particularly important in the study of tumor initiation for capturing the normal-to-cancer transition.

Next, we sought to validate the presence of this clone within the tumor and to investigate its potential role in the tumor microenvironment.

Characterization of a unique tumor clone with normal astrocyte-like phenotype

To verify the existence of clone 1 cells, we first identified gene markers unique to clone 1, including *XIST*, *RFX3*, *ADCY8*, and *GRIA1*, which can distinguish them from other tumor and normal cells (Fig. 4d, Supplementary Fig. 6a). These markers were also found in the droplet-based snRNA-seq dataset to label a putative clone 1 population, also adjacent to normal astrocytes (Supplementary Fig. 6b, c, d). As validation, we integrated the scONE-seq RNA dataset and 10X dataset to show that our scONE-seq dataset of 1210 nuclei exhaustively captured all cell types that were observed in droplet snRNA-seq of 4416 nuclei and showed that clone 1 merged with the putative clone 1 cells from the droplet-based snRNA-seq dataset (Fig. 4e).

Then, we performed histological analyses on FFPE sections from both the primary and 2R tumors of this patient to identify the cells with IDH1 (R132H) and ADCY8. Anti-IDH1 (R132H) is expected to label all tumor cells, as this is known to be a somatic mutation in IDH-mutant glioma cells^{55,57,58} (Supplementary Fig. 4a), and anti-ADCY8 is expected to mark some normal neurons and normal astrocytes in addition to clone 1 cells (Fig. 4d, Supplementary Fig. 6e). As such, putative clone 1 cells are those cells marked by double-positive staining of IDH1 (R132H) and ADCY8. First, we looked at the overall staining pattern across the whole slide section and noted that the IDH1 (R132H) positive tumor cells are distributed over the entire section for both the primary and 2R tumors. The ADCY8 signals appear stronger in 2R tumor sections and are specifically concentrated to certain regions which also express IDH (R132H) more strongly (Supplementary Fig. 7a). Interestingly, these ADCY8 positive regions are always near the IDH1 (R132H) negative 'normal adjacent' regions (Supplementary Fig. 7a). The double-positive cells that we suspect to be the putative clone 1 cells appear to be near other normal and malignant cells (Fig. 5a). These histological immune staining results provided additional details on the spatial distribution of putative clone 1 cells in the tumor sections.

The presence of these clone 1 tumor cells with their normal-like transcriptional signature among normal astrocytes and neurons prompted us to examine the gene expression of clone 1 cells associated with signaling and cell-cell communication. Several studies have demonstrated that glioma cells can form synaptic structures with normal neurons as a signaling conduit within the tumor⁵⁹⁻⁶³. Specifically, this was found to occur via tumor microtubules displaying AMPA receptors (AMPA), a glutamate receptor subtype^{62,63}. AMPARs are tetrameric, and there are four subunit proteins involved, Glut1-4, encoded by the genes *GRIA1-4* respectively^{64,65}. Interestingly, we found the *GRIA1* genes to be differentially expressed between the different tumor clones in our sample (Fig. 5b). The major clone, clone 3, expressed *GRIA2-4* and does not express *GRIA1*; clone 1, however, is the only tumor subpopulation that expresses *GRIA1* and all three other *GRIA* family genes are expressed at much lower levels. The gene expression feature of clone 1 generally resembles astrocytes, including the expression of the astrocytic marker *APOE* (Supplementary Fig. 7b), but normal astrocytes do not express *GRIA1* (Fig. 5b, Supplementary Fig. 7b). Next, we also performed ligand-receptor analysis for the different subpopulations and found TGF β signaling transcripts to be strongly and specifically expressed in normal astrocytes, clone 1 cells, and TAMs (Supplementary Fig. 7c), with clone 1 cells expressing the ligand and predominantly the TAMs expressing the receptor (Fig. 5c). These observations hint at the potential role of clone 1 cells within the tumor, and future studies expanded to additional patient samples will be necessary to determine the generalizability of these phenomena.

Discussion

We have developed the versatile scONE-seq method, which is comparable with previous methods in multiple aspects including sensitivity and accuracy, and surpasses previous methods on throughput, versatility, and sample applicability. Most importantly, we designed this technology to be compatible with frozen tissue samples that have been stored for years. This feature makes it easier to plan and perform larger-scale clinical multi-omic single-cell studies in two ways: First, by enabling studies on existing biobanked samples, which we have demonstrated herein; Second, for studies on new samples, it also removes the burden of having to immediately process tissues from clinical researchers whose priority is patient care. The application of scONE-seq on frozen tumor tissue in our study demonstrates a comprehensive investigation of the genotypic and molecular phenotypic heterogeneity of astrocytoma. By superimposing the multi-omic data, we directly observed and characterized the differentiated tumor clones, which supports the idea that tumor clones can produce a differentiation hierarchy^{7,66,67}. In addition, we found a unique tumor clone in this patient's 2R astrocytoma with a very similar transcriptomic phenotype as normal astrocytes. Putative clone 1 cells expressing the same gene markers as clone 1 were found using both independent 10X Genomics snRNA-seq as well as immunostaining on tissue sections. Incorporating the histological and transcriptomic information of clone 1 cells, we propose some potential roles of these cells in the TME, including cell-cell communication via both signaling molecules and potential tumor-neuron synapse formation. Specifically, we found clone 1 to be the only tumor subpopulation that expresses *GRIA1*, an AMPA receptor encoding gene, while all the other tumor clones only express other *GRIA* genes. *GRIA1*-encoded GluA1 subunit often forms GluA1 homomer AMPARs, which are calcium-permeable and broadly found in synapses in early development⁶⁵. Calcium-permeable AMPARs are a key signaling molecule in the tumor-neuron synapse, and the maintenance of long-term potentiation is also known to be modulated through post-translational modification of GluA1, making GluA1 critical for neural plasticity in the brain^{65,68}. The unique, differential expression of *GRIA1* in clone 1 hints at a unique, possibly multi-faceted role for clone 1 in cell-cell communication in the tumor microenvironment^{7,62,63}. We also found TGF β signaling expression to be prominently expressed in normal astrocytes, clone 1 cells, and TAMs. TGF β signaling in glioblastomas can promote invasiveness, angiogenesis, and plays a role in the suppression of the immune system; it is a therapeutic target for the treatment of gliomas. The expression of TGF β pathway-associated genes by clone 1 cells is comparable to that of the normal astrocytes within the tumor, raising intriguing questions about the role of clone 1 in tumor immune modulation, and its possible interactions with TAMs. These features of the tumor microenvironment and whether clone 1 cells are a commonly occurring cell-type in gliomas or specifically astrocytomas will be an interesting avenue of future investigation on a larger sample set. Collectively from these results, we posit that scRNA-seq-only based cancer studies could underestimate important layers of tumor heterogeneity, and that simultaneous direct DNA measurement could contribute meaningful and informative insight on tumor evolution. Meanwhile, the clonal analysis based on scWGS-only data also ignores the complex interactions within a tumor microenvironment. By deciphering the genetic and phenotypic heterogeneity within the tumor ecosystem with scONE-seq, we can reveal the interplays of clonal expansion, tumor cells differentiation hierarchy, and TME. We anticipate that future large-scale studies on more samples will be even more informative, and has the potential to inform our knowledge of cancer and therapeutics in a more generalizable manner.

Initially, scONE-seq suffered cross-contamination from DNA to RNA due to mislabeling of barcodes; this problem was resolved by optimizing the reverse transcription temperature and amplification extension time (Supplementary Table 1). Further, we checked the mapping region of the RNA data from single cells with DNA and RNA co-profiles and single cells with only RNA profiles. The exon mapping proportion difference is inconsequential (Supplementary Fig. 1e), suggesting the mislabeled molecules are a negligible fraction of the total sequenced reads. Further benchmarking and validation experiments with different sample types demonstrated that scONE-seq data correctly identifies cell types and clones simultaneously.

Benefitting from its one-tube reaction system, scONE-seq possesses very high scalability, and a higher throughput version can be achieved by simply adapting this method to a robotic liquid handler. Alternatively, producing scONE-seq and droplet-based single-cell data in parallel and then integrating them, is also a useful complementary, multi-omics approach to study cancer with high throughput. Moreover, additional processing can be easily added to the scONE-seq workflow to enable profiling of more layers of information: To detect chromatin accessibility simultaneously, an additional nuclei tagmentation step⁶⁹⁻⁷¹ with customized ATAC adaptors could be added before FACS sorting; Similarly, quantitative protein estimation⁷² could be achieved by using DNA-barcoded antibodies before single-cell sorting steps of the scONE-seq; and by jointly performing whole-exome capturing or any hybridized target sequencing panels with the standard scONE-seq library, paired high-depth single-cell somatic mutation information could also be integrated into the scONE-seq dataset. In general, scONE-seq has very strong scalability to measure different signals from a cell. We expect that scONE-seq will be a powerful tool to dissect cellular heterogeneity and will inspire other ultra-high-throughput single-cell multi-omics methods development.

Methods

Single cell or nucleus isolation. HCT116 (ATCC), NPC43 (kindly provided by Prof. KW Lo in Chinese University of Hong Kong), HUVEC (LONZA), and H9 (WiCell) cells are dissociated with trypsin-EDTA (0.25%) solution (Thermo Fisher) and stained with propidium iodide (1mg/ml, Thermo Fisher) to exclude dead cells.

Fresh whole blood was taken in the HKUST clinic from a healthy human donor. Lymphocytes were isolated via Ficoll-Paque PLUS (GE Healthcare) density centrifugation. The red blood cells were removed with 1X Red Blood Cell lysis buffer (Thermo Fisher).

The frozen IDH1-mutant astrocytoma tissue (stored at -80 °C) was obtained from Prince of Wales Hospital. The nuclei isolation protocol is based on previous studies^{73,74}. In brief, the homogenization method was used to prepare nuclei. 100mg frozen tissue was put into pre-chilled glass douncer contained 2ml of 1X homogenization buffer (5mM CaCl₂, 3mM Mg(Ac)₂, 10mM Tris, 167 μM β-mercaptoethanol, 320 mM sucrose, 0.1 mM EDTA, 0.1% NP40, 1U/ml RNase inhibitor, pH=7.8). The 'loose' and 'tight' pestles were applied to homogenize the frozen tissue. The homogenized suspension was then filtered with a 40 μm cell strainer (Corning), and nuclei can be spun down at 1000 g, 10 mins, 4°C. We utilized Sucrose Density Gradient to remove cell debris. Nuclei were resuspended in 3.0 ml low sucrose buffer (320 mM sucrose, 10 mM HEPES, 5 mM CaCl₂, 3mM Mg(Ac)₂, 0.1 mM EDTA, 1mM DTT, 0.2 U/ml RNase inhibitor, pH=8.0). We then layered 12.5 ml of density sucrose buffer (1 M sucrose, 10 mM HEPES, 3mM Mg(Ac)₂, 1mM DTT, pH=8.0) underneath the low sucrose buffer homogenate to centrifuge at 3200g for 20 min at 4°C. After removing the supernatant, nuclei were resuspended in sorting buffer (1XPBS, 1%BSA, 1U/ml RNase inhibitor, 0.1 mg/ml DAPI (Thermo Fisher)).

Cells or nuclei were then loaded to Aria III flow cytometer (BD Biosciences) to sort single cells into PCR tubes (96 or 384 PCR plates) containing lysis buffer. The lysis buffer consisted of 2.5 U/μl RNase Inhibitor (NEB), 0.15% Triton X-100 (Sigma), and 6 μM DTT (Thermo Fisher). The sorted sample can be stored at -80 °C for months.

Generation of scONE-seq libraries. To start the scONE-seq pre-amplification, the proteinase K (Sigma) was used to completely lyse cells or nuclei. Tagmentation reaction was performed to fragment the genome DNA and add the DNA-specific barcode. This reaction includes the following components, 6 mM MgCl₂, 0.5 mM dNTP (NEB), 8.5 mM TAPs-NaOH, 1.5 U/μl RNase Inhibitor, 0.05 U KAPA polymerase (Roche), 8 % PEG8000, and One-Tn5^{35,36} with customized adaptor (GTCTCGTGGGCTCGGTCATGN₅AGATGTGTATAAGAGA CAG) (Novoprotein). The reaction was incubated at 55°C for 10 mins followed by 72 °C for 10 mins. Then, proteinase K or thermolabile proteinase K (NEB) was used to deactivate the enzyme in the buffer. Thereafter, we performed reverse transcription with the following components, 40 U SuperScript™ III Reverse Transcriptase (Thermo Fisher), 70 mM Tris-HCl, 1.5 U/μl RNase Inhibitor, 8 mM MgCl₂, 7 μM DTT and 0.15 μM RT primers (GTCTCGTGGGCTCGGATCGTN₅T₂₀VN; GTCTCGTGGGCTCGGATCGTN₅GGGHN; GTCTCGTGGGCTCGGATCGTN₅TTTTVN). Reverse transcription was carried out at 12 °C for 12 sec followed by gradient increasing to 50 °C, and incubating for 50 min, and 55 °C for 50 min. Subsequently, the residual primers and RNA were removed with thermolabile EXO I (NEB), RNase If (NEB), and RNase H (NEB). Then, the terminal transferase (NEB) was used to add the C-tail to cDNA fragments. This reaction was performed at 37°C for 5 mins and the enzyme was immediately deactivated with thermolabile proteinase K. Second strand synthesis was then performed by adding, 0.3 μM 3' adaptor (GTCTCGTGGGCTCGGATCGTN₅GGGHN), 1 μl KAPA HiFi Fidelity Buffer (5X), 0.7 mM (NH₄)₂SO₄, and 0.1 μl KAPA Polymerase. The reaction was incubated at 72 °C for 5 min; 10 cycles of (1 min at 48 °C; 1 min at 72 °C); and 5 min at 72 °C, in a thermal cycler. Additional residual primers removal reaction was performed with Exo I (NEB). Lastly, 14 μl KAPA HotStart ReadyMix (2X), 1.5 mM (NH₄)₂SO₄, 2% DMSO (Thermo Fisher), 1.2 μM amplification primer (GATGTGTGGAGGTCTCGTGGGCTCGG) was added to amplify DNA and RNA simultaneously. The PCR was performed at 98 °C for 4 min; 18-20 cycles of (20 s at 98 °C; 4.25 min at 72 °C); and 10 min at 72 °C, in a thermal cycler. All primer sequences can be found in Supplementary Table 2.

scONE-seq sequencing libraries construction. Pre-amplified samples were purified with Ampure XP beads (Beckman). Samples were diluted to 0.1 ng/μl and performed tagmentation reaction with the following components, 1X TAPs buffer (50 mM TAPS-NaOH, 25 mM MgCl₂, pH = 8.0), 8 % PEG8000, 0.001 μl P5-Tn5 (Adaptor sequence: TCGTCGGCAGGTCAGATGTGT ATAAGAGACAG). The reaction was performed at 55°C for 15 min. Samples were then amplified with Illumina sequencing index primers (Supplementary Table 2) (Sangon) by using KAPA HiFi HotStart Polymerase Kit (Roche). The enrichment PCR was incubated at 95 °C for 10 min; 10-11 cycles of (20 s at 98 °C; 15 s at 60 °C; 30 s at 72 °C); and 2 min at 72 °C, in a thermal cycler. Samples were then pooled and purified with Ampure XP beads. scDASH protocol was then used to remove the abundant ribosome and mitochondrial sequences^{75,76}. Double size

selection can be performed to optimize the library size. The library was then sequenced on Illumina NextSeq500 with customized sequencing primers (Supplementary Table 2).

DNA and RNA data separation. Sequencing data were firstly filtered with fastp⁷⁷. Fastq files were then separated into DNA fastq files, RNA fastq files, and Unmatched fastq files with seqkit, seqtk, and bbdup^{78–80}. During this process, UMI of the reads was extracted and labeled to fastq files head with fastp⁷⁷.

DNA data analysis. DNA fastq files were mapped to hg38 (ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000/001/405/GCA_000001405.15_GRCh38/seqs_for_alignment_pipelines.ucsc_ids/GCA_000001405.15_GRCh38_no_alt_analysis_set.fna) with BWA mem⁸¹. To perform UMI-based deduplication, read2 reads in bam files were extracted with samtools⁸² and deduplicated with umi_tools⁸³. The deduplicated read2 reads were used to extract its paired read1 and these paired fastq were then re-aligned to hg38 with BWA mem^{81,84}.

If only performing the counts-based copy number variation analysis, Ginkgo was used to generate the normalized counts⁴³. If performing the allele-specific copy number variation analysis, CHISEL was used to generate allele frequency information^{85,86}. The integer copy number calculation was based on the previous studies^{21,22,87}. In this pipeline, the segmentation was performed with copynumber and aCGH in R⁸⁸.

WES data from different tumor samples were analyzed with CNVkit and THetA to infer the copy number profiles^{89–91}. The integer copy number data from WES and scONE-seq were combined to perform the phylogenetic reconstruction, adopted from a previous method^{21,92,93}.

RNA data analysis. UMI-based deduplication was also performed with RNA fastq files. The workflow kept the same except replacing the BWA with STAR⁹⁴. Then, the fastq files can be quantified with Kallisto⁹⁵ (cDNA quantification) or Salmon⁹⁶ (pre-mature RNA quantification). 10x snRNA-seq data was quantified with kb-python⁹⁷. The expression data were analyzed using Seurat with sctransform pipeline (normalization, dimension reduction, dataset integration, finding clusters, differential gene analysis)^{98–100}. The astrocytoma cellular states scoring was performed following the original paper¹³. The ligand-receptor analysis was performed with CellChat¹⁰¹.

Visualization. Plots were created using the ggplot2 and ggtrees R package^{102–104}. Heatmaps were created with the ComplexHeatmap package¹⁰⁵. Figures were prepared in Inkscape¹⁰⁶.

IHC analysis. Slides were obtained from Prince of Wales Hospital. Xylene and ethanol were used to remove wax. Antigen retrieval was performed with Sodium Citrate Buffer (Thermo Fisher) at 98°C for 15 min. Blocking was performed in 10% normal serum (goat and donkey, Abcam) with 1% BSA in PBST buffer (0.05% Triton X-100). IDH1(R132H) antibody (Dianova, 1:40) and ADCY8 antibody (Abcam, 1:200) were added to slides and incubated at 4°C overnight in a humid box. Secondary antibodies (1:400, anti-mouse, anti-rabbit, Thermo Fisher) were used to provide the fluorescent signal. Mounting buffer with DAPI (Abcam) was used to stain the nucleus and retain fluorescence. Images were taken with Zeiss Axio Scan.Z1 Slide Scanner with 20X objective (Zeiss).

Declarations

ACKNOWLEDGEMENTS

The authors thank Prof. Yanyi Huang (Peking University) and Dr. Jianbin Wang (Tsinghua University) for their on-going support and advice on this project. We also thank Prof. Tom H. Cheung (HKUST) for providing us with Tn5 enzyme that was used in a portion of the experiments, and Prof. Kwok Wai Lo (CUHK) for providing the NPC43 cell line used in the study. We thank Dr. Danny C.Y. Leung (HKUST) for his reading and feedback on the manuscript. We also want to thank Prof. Zilong Wen (HKUST), Prof. Yingying Wei (CUHK), and Prof. Can Yang (HKUST) for their discussion and advice on the data analysis.

AUTHOR CONTRIBUTIONS

LY and ARW designed the study. LY developed scONE-seq chemistry, generated libraries, performed computational analysis, performed the IHC staining, prepared figures, and wrote the manuscript text. XLW provided input to scONE-seq chemistry, generated libraries, and performed the IHC staining. QHM performed computational analysis. SSTT prepared the nuclei and 10x library. DSCL provided input to the scONE-seq chemistry. AKC, WSP, HKN, and DTC prepare patient samples and prepare the FFPE sections. JGW and ARW supervised the work. LY and ARW wrote the manuscript. All authors edited and approved the final manuscript.

COMPETING INTERESTS

ARW and LY have filed a USPTO provisional patent on scONE-seq (application number 63/093,368)

References

1. Wu, A. R., Wang, J., Streets, A. M. & Huang, Y. Single-cell transcriptional analysis. *Annual Review of Analytical Chemistry* **10**, 439–462 (2017).

2. Gawad, C., Koh, W. & Quake, S. R. Single-cell genome sequencing: Current state of the science. *Nature Reviews Genetics* **17**, 175–188 (2016).
3. Nam, A. S., Chaligne, R. & Landau, D. A. Integrating genetic and non-genetic determinants of cancer evolution by single-cell multi-omics. *Nature Reviews Genetics* **22**, 3–18 (2021).
4. Birnbaum, K. D. Power in numbers: Single-cell RNA-seq strategies to dissect complex tissues. *Annual Review of Genetics* **52**, 203–221 (2018).
5. Villani, A. C. *et al.* Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science* **356**, (2017).
6. Treutlein, B. *et al.* Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. *Nature* **509**, 371–375 (2014).
7. Venteicher, A. S. *et al.* Decoupling genetics, lineages, and microenvironment in IDH-mutant gliomas by single-cell RNA-seq. *Science* **355**, (2017).
8. Shami, A. N. *et al.* Single-Cell RNA Sequencing of Human, Macaque, and Mouse Testes Uncovers Conserved and Divergent Features of Mammalian Spermatogenesis. *Developmental Cell* **54**, (2020).
9. McGranahan, N. & Swanton, C. Clonal Heterogeneity and Tumor Evolution: Past, Present, and the Future. *Cell* vol. 168 613–628 (2017).
10. Prager, B. C., Xie, Q., Bao, S. & Rich, J. N. Cancer Stem Cells: The Architects of the Tumor Ecosystem. *Cell Stem Cell* vol. 24 41–53 (2019).
11. Kreso, A. & Dick, J. E. Evolution of the cancer stem cell model. *Cell Stem Cell* vol. 14 275–291 (2014).
12. Shaffer, S. M. *et al.* Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance. *Nature* **546**, 431–435 (2017).
13. Neftel, C. *et al.* An Integrative Model of Cellular States, Plasticity, and Genetics for Glioblastoma. *Cell* **178**, 835-849.e21 (2019).
14. Wang, L. *et al.* The phenotypes of proliferating glioblastoma cells reside on a single axis of variation. *Cancer Discovery* **9**, 1708–1719 (2019).
15. Weng, Q. *et al.* Single-Cell Transcriptomics Uncovers Glial Progenitor Diversity and Cell Fate Determinants during Development and Gliomagenesis. *Cell Stem Cell* **24**, 707-723.e8 (2019).
16. Müller, S. *et al.* Single-cell profiling of human gliomas reveals macrophage ontogeny as a basis for regional differences in macrophage activation in the tumor microenvironment. *Genome Biology* **18**, (2017).
17. Pombo Antunes, A. R. *et al.* Single-cell profiling of myeloid cells in glioblastoma across species and disease stage reveals macrophage competition and specialization. *Nature Neuroscience* **24**, 595–610 (2021).
18. Hara, T. *et al.* Interactions between cancer cells and immune cells drive transitions to mesenchymal-like states in glioblastoma. *Cancer Cell* **39**, 779-792.e11 (2021).
19. Zhang, L. *et al.* Single-Cell Analyses Inform Mechanisms of Myeloid-Targeted Therapies in Colon Cancer. *Cell* **181**, 442-459.e29 (2020).
20. Cheng, S. *et al.* A pan-cancer single-cell transcriptional atlas of tumor infiltrating myeloid cells. *Cell* **184**, 792-809.e23 (2021).
21. Minussi, D. C. *et al.* Breast tumours maintain a reservoir of subclonal diversity during expansion. *Nature* **592**, 302–308 (2021).
22. Gao, R. *et al.* Punctuated copy number evolution and clonal stasis in triple-negative breast cancer. *Nature Genetics* **48**, 1119–1130 (2016).
23. Macaulay, I. C. *et al.* G&T-seq: Parallel sequencing of single-cell genomes and transcriptomes. *Nature Methods* **12**, 519–522 (2015).
24. Dey, S. S., Kester, L., Spanjaard, B., Bienko, M. & van Oudenaarden, A. Integrated genome and transcriptome sequencing of the same cell. *Nature Biotechnology* **33**, 285–289 (2015).
25. Hou, Y. *et al.* Single-cell triple omics sequencing reveals genetic, epigenetic, and transcriptomic heterogeneity in hepatocellular carcinomas. *Cell Research* **26**, 304–319 (2016).
26. Bian, S. *et al.* *Single-cell multiomics sequencing and analyses of human colorectal cancer*. <https://www.science.org>.
27. Zachariadis, V., Cheng, H., Andrews, N. & Enge, M. A Highly Scalable Method for Joint Whole-Genome Sequencing and Gene-Expression Profiling of Single Cells. *Molecular Cell* **80**, 541-553.e5 (2020).
28. Guo, F. *et al.* Single-cell multi-omics sequencing of mouse early embryos and embryonic stem cells. *Cell Research* **27**, 967–988 (2017).
29. Oguchi, Y., Ozaki, Y., Abdelmoez, M. N. & Shintaku, H. NanoSINC-seq dissects the isoform diversity in subcellular compartments of single cells. *Science Advances* **7**, (2021).
30. Abdelmoez, M. N. *et al.* SINC-seq: Correlation of transient gene expressions between nucleus and cytoplasm reflects single-cell physiology. *Genome Biology* **19**, (2018).
31. Kivioja, T. *et al.* Counting absolute numbers of molecules using unique molecular identifiers. *Nature Methods* **9**, 72–74 (2012).

32. Klein, A. M. *et al.* Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* **161**, 1187–1201 (2015).
33. Macosko, E. Z. *et al.* Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214 (2015).
34. Reznikoff, W. S. Transposon Tn5. *Annual Review of Genetics* vol. 42 269–286 (2008).
35. Picelli, S. *et al.* Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Research* **24**, 2033–2040 (2014).
36. Hennig, B. P. *et al.* Large-scale low-cost NGS library preparation using a robust Tn5 purification and tagmentation protocol. *G3: Genes, Genomes, Genetics* **8**, 79–89 (2018).
37. Sheng, K., Cao, W., Niu, Y., Deng, Q. & Zong, C. Effective detection of variation in single-cell transcriptomes using MATQ-seq. *Nature Methods* **14**, 267–270 (2017).
38. Tang, F. *et al.* mRNA-Seq whole-transcriptome analysis of a single cell. *Nature Methods* **6**, 377–382 (2009).
39. Sheng, K., Cao, W., Niu, Y., Deng, Q. & Zong, C. Effective detection of variation in single-cell transcriptomes using MATQ-seq. *Nature Methods* **14**, 267–270 (2017).
40. Picelli, S. *et al.* Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nature methods* **10**, 1096–8 (2013).
41. Picelli, S. *et al.* Full-length RNA-seq from single cells using Smart-seq2. *Nature protocols* **9**, 171–81 (2014).
42. Fan, X. *et al.* Single-cell RNA-seq transcriptome analysis of linear and circular RNAs in mouse preimplantation embryos. *Genome Biology* **16**, (2015).
43. Garvin, T. *et al.* Interactive analysis and assessment of single-cell copy-number variations. *Nature Methods* vol. 12 1058–1060 (2015).
44. Sun, J. C. & Lanier, L. L. NK cell development, homeostasis and function: Parallels with CD8 + T cells. *Nature Reviews Immunology* vol. 11 645–657 (2011).
45. Farber, D. L., Yudanin, N. A. & Restifo, N. P. Human memory T cells: Generation, compartmentalization and homeostasis. *Nature Reviews Immunology* vol. 14 24–35 (2014).
46. Pizzolato, G. *et al.* Single-cell RNA sequencing unveils the shared and the distinct cytotoxic hallmarks of human TCRV δ 1 and TCRV δ 2 $\gamma\delta$ T lymphocytes. *Proceedings of the National Academy of Sciences of the United States of America* **116**, 11906–11915 (2019).
47. Uhlen, M. *et al.* A genome-wide transcriptomic analysis of protein-coding genes in human blood cells. *Science* **366**, (2019).
48. Lin, W. *et al.* Establishment and characterization of new tumor xenografts and cancer cell lines from EBV-positive nasopharyngeal carcinoma. *Nature Communications* **9**, (2018).
49. Ben-David, U. *et al.* Genetic and transcriptional evolution alters cancer cell line drug response. *Nature* **560**, 325–330 (2018).
50. Ceccarelli, M. *et al.* Molecular Profiling Reveals Biologically Discrete Subsets and Pathways of Progression in Diffuse Glioma. *Cell* **164**, 550–563 (2016).
51. Hu, H. *et al.* Mutational Landscape of Secondary Glioblastoma Guides MET-Targeted Trial in Brain Tumor. *Cell* **175**, 1665-1678.e18 (2018).
52. Habib, N. *et al.* Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nature Methods* **14**, 955–958 (2017).
53. Mu, Q. *et al.* MYC amplification at diagnosis drives therapy-induced hypermutation of recurrent glioma. *ResearchSquare* (2021).
54. Sanchez-Vega, F. *et al.* Oncogenic Signaling Pathways in The Cancer Genome Atlas. *Cell* **173**, 321-337.e10 (2018).
55. Wang, J. *et al.* Clonal evolution of glioblastoma under therapy. *Nature Genetics* **48**, 768–776 (2016).
56. Louis, D. N. *et al.* The 2021 WHO Classification of Tumors of the Central Nervous System: a summary. *Neuro-Oncology* **23**, (2021).
57. Johnson, B. E. *et al.* Mutational analysis reveals the origin and therapy-driven evolution of recurrent glioma. *Science* **343**, (2014).
58. Barthel, F. P. *et al.* Longitudinal molecular trajectories of diffuse glioma in adults. *Nature* **576**, (2019).
59. Venkataramani, V., Tanev, D. I., Kuner, T., Wick, W. & Winkler, F. Synaptic input to brain tumors: clinical implications. *Neuro-oncology* **23**, 23–33 (2021).
60. Jung, E. *et al.* Emerging intersections between neuroscience and glioma biology. *Nature Neuroscience* vol. 22 1951–1960 (2019).
61. Venkatesh, H. S. *et al.* Electrical and synaptic integration of glioma into neural circuits. *Nature* **573**, 539–545 (2019).
62. Venkataramani, V. *et al.* Glutamatergic synaptic input to glioma cells drives brain tumour progression. *Nature* **573**, 532–538 (2019).
63. Osswald, M. *et al.* Brain tumour cells interconnect to a functional and resistant network. *Nature* **528**, 93–98 (2015).

64. Henley, J. M. & Wilkinson, K. A. Synaptic AMPA receptor composition in development, plasticity and disease. *Nature Reviews Neuroscience* vol. 17 337–350 (2016).
65. Diering, G. H. & Huganir, R. L. The AMPA Receptor Code of Synaptic Plasticity. *Neuron* vol. 100 314–329 (2018).
66. Tirosh, I. *et al.* Single-cell RNA-seq supports a developmental hierarchy in human oligodendroglioma. *Nature* **539**, 309–313 (2016).
67. Neftel, C. *et al.* An Integrative Model of Cellular States, Plasticity, and Genetics for Glioblastoma. *Cell* **178**, 835–849.e21 (2019).
68. Henley, J. M. & Wilkinson, K. A. Synaptic AMPA receptor composition in development, plasticity and disease. *Nature Reviews Neuroscience* **2016** 17:6 17, 337–350 (2016).
69. Buenrostro, J. D. *et al.* Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523**, 486–490 (2015).
70. Cao, J. *et al.* SINGLE-CELL GENOMICS Joint profiling of chromatin accessibility and gene expression in thousands of single cells. *Science* vol. 361 (2018).
71. Lareau, C. A. *et al.* Droplet-based combinatorial indexing for massive-scale single-cell chromatin accessibility. *Nature Biotechnology* **37**, 916–924 (2019).
72. Stoekius, M. *et al.* Simultaneous epitope and transcriptome measurement in single cells. *Nature Methods* **14**, 865–868 (2017).
73. Corces, M. R. *et al.* An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nature Methods* **14**, 959–962 (2017).
74. Matson, K. J. E. *et al.* Isolation of adult spinal cord nuclei for massively parallel single-nucleus RNA sequencing. *Journal of Visualized Experiments* **2018**, (2018).
75. Gu, W. *et al.* Depletion of Abundant Sequences by Hybridization (DASH): Using Cas9 to remove unwanted high-abundance species in sequencing libraries and molecular counting applications. *Genome Biology* **17**, (2016).
76. Loi, D. S. C., Yu, L. & Wu, A. R. Effective ribosomal RNA depletion for single-cell total RNA-seq by scDASH. *PeerJ* **9**, e10717 (2021).
77. Chen, S., Zhou, Y., Chen, Y. & Gu, J. Fastp: An ultra-fast all-in-one FASTQ preprocessor. in *Bioinformatics* vol. 34 (2018).
78. Shen, W., Le, S., Li, Y. & Hu, F. SeqKit: A cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLoS ONE* **11**, (2016).
79. Li, H. seqtk Toolkit for processing sequences in FASTA/Q formats. *GitHub* **767**, (2012).
80. Bushnell, B., Rood, J. & Singer, E. BBTools Software Package. *PLoS ONE* vol. 12 e0185056 <https://sourceforge.net/projects/bbmap/> (2017).
81. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, (2009).
82. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, (2009).
83. Smith, T., Heger, A. & Sudbery, I. UMI-tools: Modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Research* **27**, (2017).
84. Quinlan, A. R. & Hall, I. M. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, (2010).
85. Zaccaria, S. & Raphael, B. J. Characterizing allele- and haplotype-specific copy numbers in single cells with CHISEL. *Nature Biotechnology* **39**, 207–214 (2021).
86. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nature Genetics* **48**, (2016).
87. Navin, N. *et al.* Tumour evolution inferred by single-cell sequencing. *Nature* **472**, 90–95 (2011).
88. Nielsen, G. *et al.* Copynumber: Efficient algorithms for single- and multi-track copy number segmentation. *BMC Genomics* **13**, (2012).
89. Talevich, E., Shain, A. H., Botton, T. & Bastian, B. C. CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing. *PLoS Computational Biology* **12**, (2016).
90. Oesper, L., Mahmood, A. & Raphael, B. J. THetA: Inferring intra-tumor heterogeneity from high-throughput DNA sequencing data. *Genome Biology* **14**, (2013).
91. Oesper, L., Satas, G. & Raphael, B. J. Quantifying tumor heterogeneity in whole-genome and whole-exome sequencing data. *Bioinformatics* **30**, (2014).
92. Paradis, E. & Schliep, K. Ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**, (2019).

93. Desper, R. & Gascuel, O. Fast and accurate phylogeny reconstruction algorithms based on the minimum-evolution principle. in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* vol. 2452 (2002).
94. Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, (2013).
95. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology* **34**, (2016).
96. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods* **14**, (2017).
97. Melsted, P. *et al.* Modular, efficient and constant-memory single-cell RNA-seq preprocessing. *Nature Biotechnology* **39**, (2021).
98. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nature Biotechnology* **36**, (2018).
99. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902.e21 (2019).
100. Finak, G. *et al.* MAST: A flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biology* **16**, (2015).
101. Jin, S. *et al.* Inference and analysis of cell-cell communication using CellChat. *Nature Communications* **2021 12:1 12**, 1–20 (2021).
102. Gómez-Rubio, V. ggplot2 - Elegant Graphics for Data Analysis (2nd Edition) . *Journal of Statistical Software* **77**, (2017).
103. Kassambara, A. ggpubr: "ggplot2" Based Publication Ready Plots. *R package version 0.4.0* (2020).
104. Yu, G., Smith, D. K., Zhu, H., Guan, Y. & Lam, T. T. Y. ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution* **8**, (2017).
105. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, (2016).
106. Inkscape.org. Draw Freely | Inkscape. *inkscape.org* (2020).

Figures

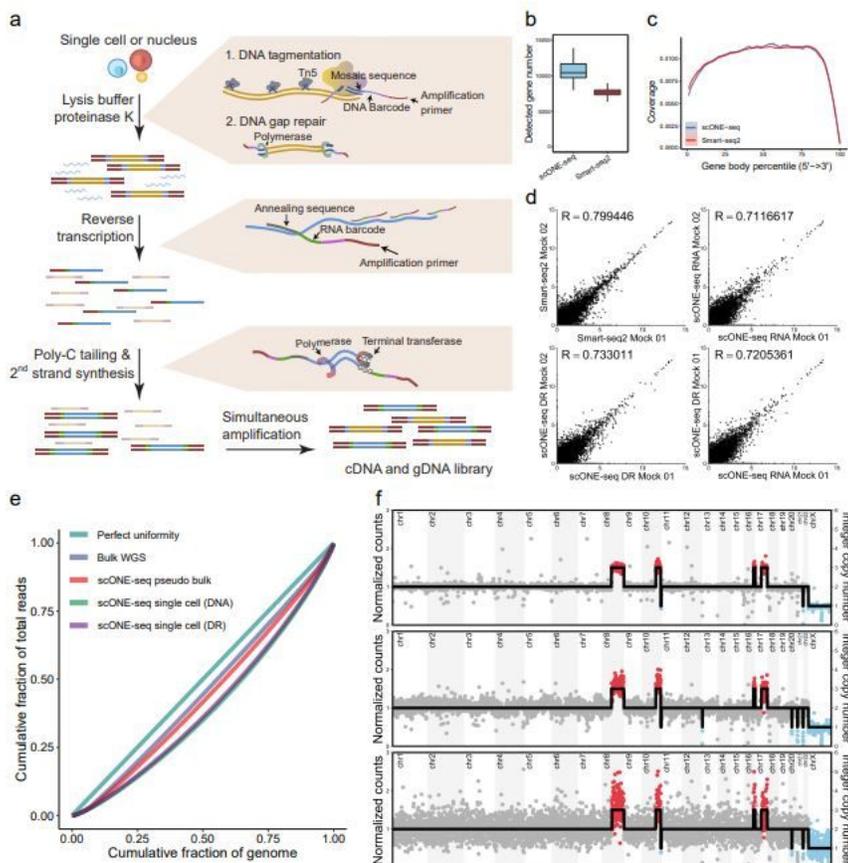


Figure 1

Overview of scONE-seq: schematic and benchmark. a, Workflow of scONE-seq. DNA and RNA barcodes are added by Tn5 with custom adaptor and RT primers respectively. P7-Illumina sequencing primer is included in this pre-amplification process. P5-Illumina sequencing primer is added in the later library construction step with a second Tn5 tagmentation custom adaptor. b, Gene detection sensitivity for scONE-seq (n = 86, HCT116 cells) and Smart-seq2 (n = 94, HCT116 cells), downsampled to similar sequencing depth (0.2 million mapped reads) ($P < 2 \times 10^{-16}$, t-test). Shown are numbers of genes detected over 1 count. c, Gene body coverage for scONE-seq (n = 86) and Smart-seq2 (n = 94). Cell-cell variations are shown with error areas. d, Accuracy across mock samples for scONE-seq (RNA control, DNA+RNA group) and Smart-seq2. Pearson correlations were calculated from log-transformed TPM values. e, The Lorenz curve of bulk and scONE-seq single-cell data. Percentiles of the genome covered are plot against the cumulative fraction of reads. A perfect coverage uniformity results in a straight line with slope 1. f, g, h, Dots plot with normalized counts across the genome and solid line plots for corresponding estimated integer copy number. Amplification regions are highlighted with red; deletion regions are highlighted with light blue. Data from the bulk HCT116 whole genome sequencing (e), the HCT116 scONE-seq pseudo-bulk data (f, n = 86), and a single-cell HCT116 scONE-seq data (g).

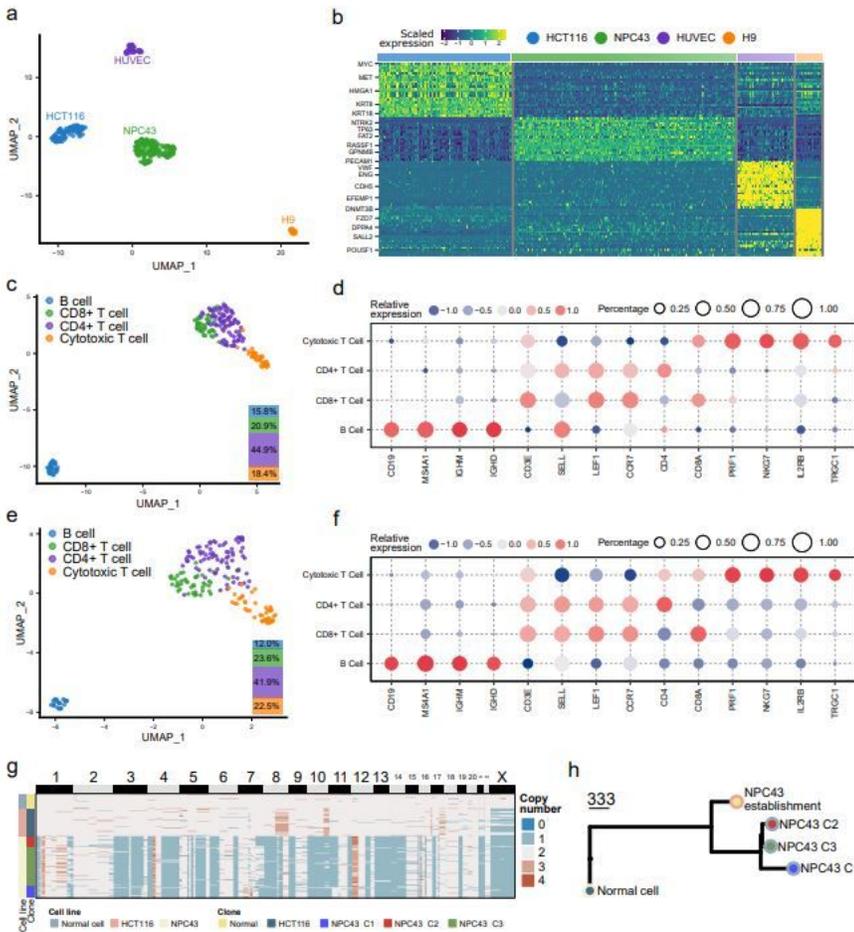


Figure 2

scONE-seq cell-types classification and CNAs clone identification. a, UMAP of scONE-seq cell lines RNA data, cells from same the cell line are clustered together. b, Differential expression genes (DEGs) heatmap; DEGs separate the cells based on their cell type; common markers for these cell lines are labeled in the heatmap. c, e, UMAP of lymphocytes RNA data from scONE-seq (c) and Smart-seq2 (e); cell-types annotations are based on known markers of immune cells. Cell-type composition shows no difference between two datasets ($P = 0.5109$, Chi-squared test). d, f, Dot plots of markers used for cell-types annotations with the scONE-seq dataset (d) and Smart-seq2 dataset (f). g, Copy number profiles calculated with scONE-seq cell line DNA dataset; cells are organized by hierarchical clustering (normal n = 27; HCT116 n = 48; NPC43 n = 108). h, The minimum evolution tree with diploid as root; NPC43 cells used in this study acquired more CNAs compared with the genome state when the cell line was established. Unit shows the Manhattan distance.

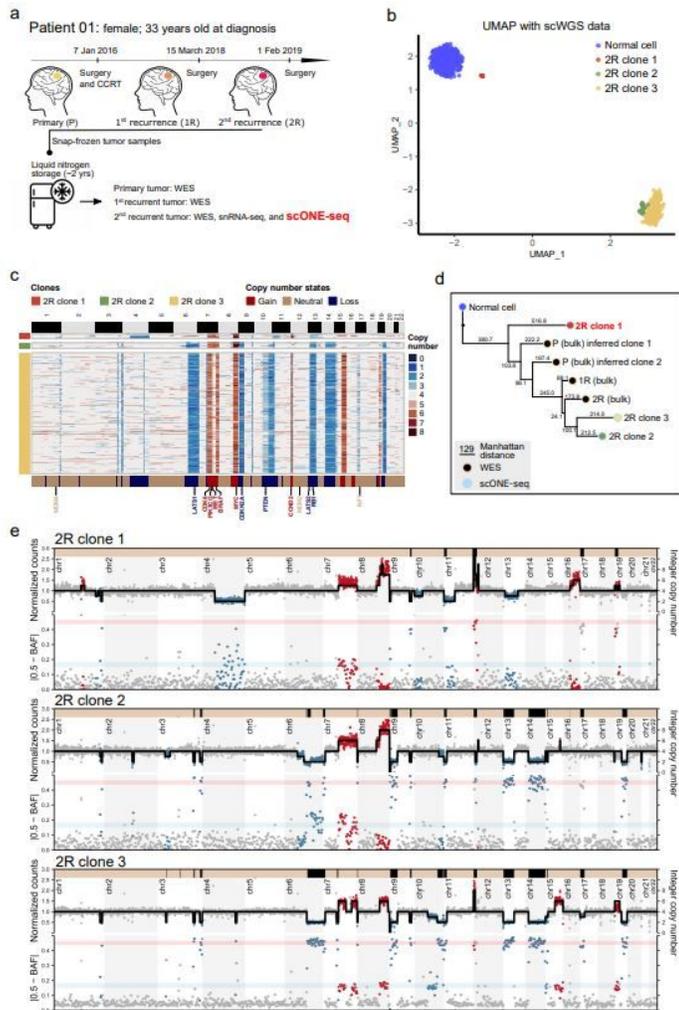


Figure 3

scONE-seq reveals the clonal composition of the IDH1-mutant astrocytoma. a, Schematic showing the patient history for the sample used in this study. The patient has been diagnosed as IDH1-mutant (Grade IV) astrocytoma. Surgery was performed to excise the tumor, and concurrent chemoradiotherapy (CCRT) was then applied. The recurrences of the tumor were excised without further drug treatment. Tumor specimens were snap-frozen in the liquid nitrogen tank and stored for two years before subject to nuclei extraction. b, UMAP of scONE-seq DNA copy number data shows the 4 genome states in this IDH1-mutant astrocytoma sample. 2R – second recurrence. c, Heatmap of integer copy numbers of all the 2R astrocytoma cells profiled; 3 clones with different copy number profiles are observed. The bottom annotation bar represents the CNAs in this tumor sample. Some commonly known glioma/astrocytoma driver genes are shown. Amplified genes are highlighted with red; deleted genes are highlighted with dark blue. d, The minimum evolution tree with diploid as root; the WES data inferred CNV from the same patient were integrated to show the evolutionary relationship between the tumor recurrences and their various clones. The P clones are inferred from bulk WES data (see Methods). P – primary; 1R – first recurrence. e, The top panel shows the dots plot with normalized counts across the human genome and the solid line representing the estimated integer copy number. The bottom panel shows the mirrored BAF dots plot across the genome. Relatively amplified regions are highlighted with red; relatively deleted regions are highlighted with blue. If dots are close to the red belt in the mirrored BAF dots plot, this indicates that there are LOH in those regions. If dots are close to the blue belt in the mirrored BAF dots plot, this indicates that there are imbalanced haplotypes in that regions. The top bar highlights the LOH regions of the genome. The clonal pseudo-bulk genome information for each 2R clone is also shown.

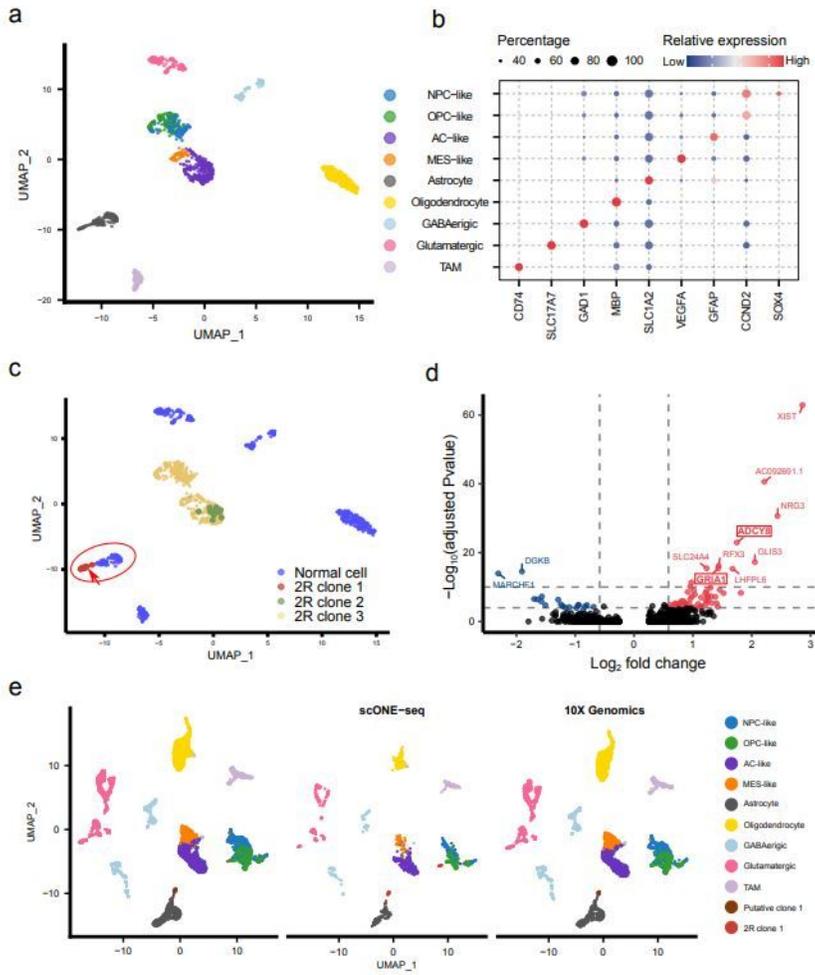


Figure 4

scONE-seq reveals the tumor microenvironment composition of an IDH1-mutant astrocytoma. a, UMAP of scONE-seq RNA data shows the TME composition cell types in this second recurrence IDH1-mutant astrocytoma sample. Tumor cells are classified into 4 cellular states based on their meta-module scores. b, Dot plot shows some markers for the annotation of cell types. c, UMAP of scONE-seq RNA data annotated with clonal information. The 2R clone 1 cells are clustered with normal astrocytes. d, Volcano plot shows the DEGs between the 2R clone 1 and clone 3 cells. Genes with higher expression in clone 1 cells are colored with red. Genes with higher expression in clone 3 are colored with blue. e, UMAP showing the integration of scONE-seq and 10x snRNA-seq dataset. The integrated data (left) retains all cell types identified with scONE-seq or 10x snRNA-seq. The split-UMAP (right) shows adequate mixing of cell-types found separately by the 2 methods, with clone 1 cells (from scONE-seq) and putative clone 1 cells (from snRNA-seq) falling into the same integrated cluster.

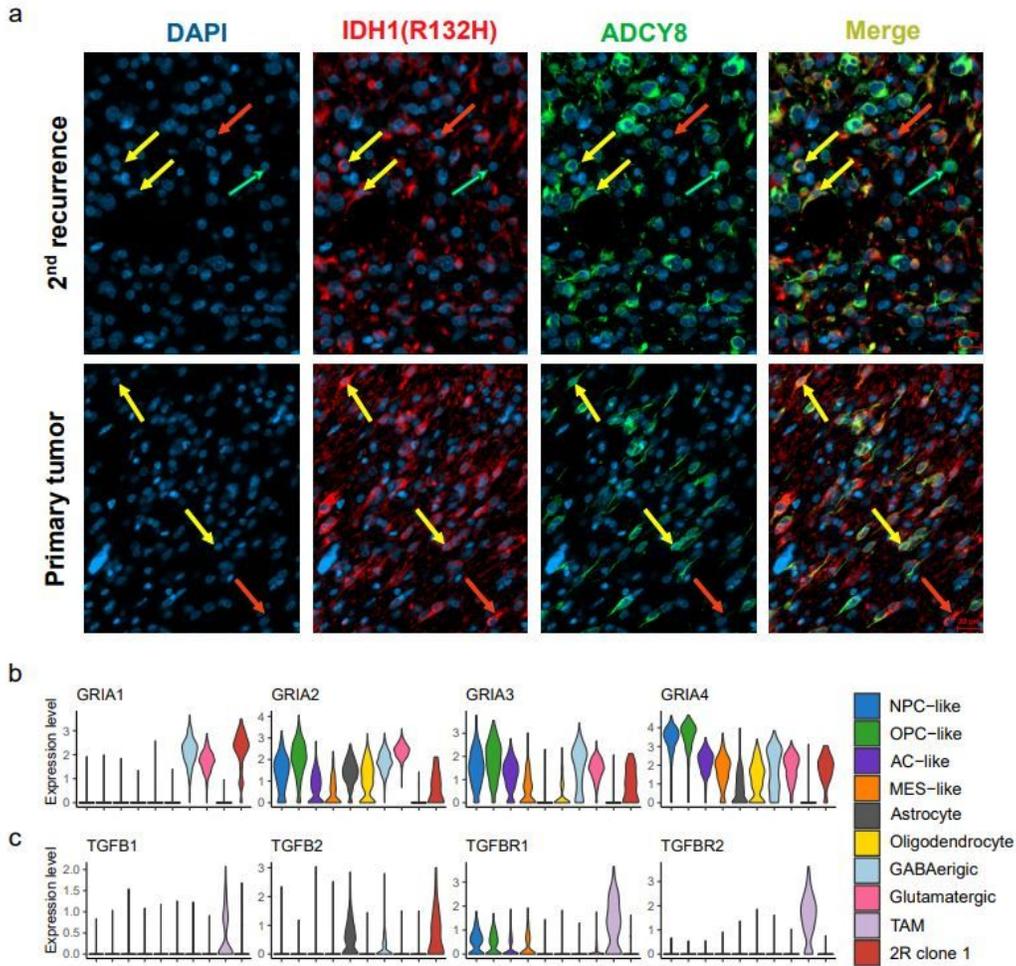


Figure 5

Characterization of 2R clone 1 features. a, The immunofluorescent images co-label the IDH1 (R132H) and ADCY8 in the FFPE section of the patient. The top panel shows images from the 2R tumor; the bottom shows images from the primary tumor from the same patient. The yellow arrow indicates the co-stained putative clone 1 cells; the red arrow indicates the other tumor cells; the Green arrow indicates normal astrocytes or GABAergic neurons. b, The AMPAR subunit-encoding gene expression pattern in this 2R IDH1-mutant astrocytoma. GRIA1 is highly expressed in clone 1 cells. c, The TGF β signaling genes expression pattern in this 2R IDH1-mutant astrocytoma. TGFB2 is highly expressed in clone 1 cells and normal astrocytes. The receptors are mostly expressed in TAMs.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryTable1.xlsx](#)
- [SupplementaryTable2.xlsx](#)
- [20211005supplementaryfigures.pdf](#)