



Article

De Novo Detection of Clonal Structure and Evolution in Single-Cell and Spatial Transcriptomes

Shihao Bai [†], Xianbin Su ^{*,†} , Ziyao Chen and Ze-Guang Han ^{*} 

Key Laboratory of Systems Biomedicine (Ministry of Education), Shanghai Center for Systems Biomedicine, Shanghai Jiao Tong University, Shanghai 200240, China; b369259046@sjtu.edu.cn (S.B.); unicornczy@sjtu.edu.cn (Z.C.)

^{*} Correspondence: xbsu@sjtu.edu.cn (X.S.); hanzg@sjtu.edu.cn (Z.-G.H.)

[†] These authors contributed equally to this work.

Abstract

Tumors are composed of cellular populations with distinct genotypes and phenotypes, which dynamically evolve over time and during treatment. This process is known as clonal evolution, and it is difficult to reveal fine-scale clonal structure with traditional bulk sequencing. Although single-cell genome sequencing could enable reconstruction of tumor clonal evolution, it remains technically challenging and the number of single cells profiled is generally insufficient due to high cost. To address this issue, we developed scClone, a computational toolkit that integrates variant detection and genotype inference for single-cell RNA-seq (scRNA-seq) and spatial transcriptomic data. It further provides interactive visualization of clonal structure and dynamic evolution. scClone addresses key limitations inherent to scRNA-seq, such as expression drop-out and allelic imbalance, and incorporates cell type or state annotation with mutational signature analysis to enable comprehensive profiling of tumor heterogeneity. scClone demonstrated robust performance across multiple datasets—generated from both full-length and fragmented RNA sequencing—by accurately reproducing mutation profiles and resolving clonal mixtures in myeloma, hepatocellular carcinoma and pancreatic cancer. Additionally, scClone has been applied to spatial transcriptomics, enabling the delineation of clonal structures within histological sections from ovarian cancer and cutaneous squamous cell carcinoma. In summary, our results demonstrate that scClone can extract genetic information from scRNA-seq datasets, thereby successfully establishing genotype–phenotype associations at the single-cell level and providing insights into the clonal evolution of tumors.

Keywords: clone inference; single-cell mutation; single-cell transcriptomes; spatial transcriptomes; mutational signature



Academic Editor: Mauro Coluccia

Received: 27 October 2025

Revised: 20 November 2025

Accepted: 24 November 2025

Published: 26 November 2025

Citation: Bai, S.; Su, X.; Chen, Z.; Han, Z.-G. De Novo Detection of Clonal Structure and Evolution in Single-Cell and Spatial Transcriptomes. *Int. J. Mol. Sci.* **2025**, *26*, 11428. <https://doi.org/10.3390/ijms262311428>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The progression of cancer is closely related to Darwinian evolution. Genetic or epigenetic variations could alter the molecular phenotypes of individual cells. Tumors at the time of diagnosis typically consist of multiple cell populations with distinct genetic profiles, which are referred to as clones [1]. Clonal evolution describes the process through which tumor clones undergo phylogenetic diversification, driven by selective pressures exerted by the tumor microenvironment (TME) or therapeutic interventions, ultimately shaping the tumor's evolutionary trajectory. Clonal diversity within tumors drives tumor heterogeneity, leading to differential growth advantages, metastatic potential, therapeutic responsiveness,

and prognostic outcomes, which collectively underlie clinical challenges in cancer management. Thus, precise delineation of the clonal origin and architectural dynamics is critical for reconstructing tumor evolutionary trajectories and elucidating their clinical implications in therapeutic resistance, metastatic dissemination, and disease recurrence [2].

Methods such as fluorescence in situ hybridization (FISH) and immunohistochemistry (IHC) allow us to vaguely observe the clonal structure of tumors. In addition, techniques such as CRISPR-based lineage tracing and fluorescent labeling methods [3], which track cell division and proliferation during tumor progression, are striving to reveal the process of tumor clonal evolution. The advent of high-throughput DNA sequencing has enabled the application of bioinformatics pipelines to deconvolve hidden evolutionary trajectories from bulk genomic sequencing data, primarily by inferring clonal architectures through variant allele frequency (VAF)-based clustering and phylogenetic reconstruction. However, current strategies for bulk-level clonal evolution analysis suffer from an inherent flaw: they use the clustering of mutations to infer tumor clones, but the essence of tumor clones is the clustering of cell lineages [4]. For example, a subclone is defined as a genetically distinct group of tumor cells resulting from a clonal expansion; however, in practice, it is often detected by a group of mutations whose variant allele frequency (VAF) equals approximately 0.25 in bulk sequencing. Inferring the complex clonal structure within tumors using one-dimensional bulk VAF values inherently introduces inaccuracies and deviates from the true results, as the nonlinear relationship between VAFs and clonal propagation dynamics (e.g., clonal sweeps, neutral drift, or selective sweeps) could lead to either false negatives (underdetection of subclones) or false positives (overestimation of clonal dominance). This suggests that clonal analysis indeed requires higher-resolution investigation.

Single-cell multi-omics provides unprecedented resolution to dissect the mechanisms underlying occurrence and progression of tumors, revealing clonal evolution trajectories, phenotypic plasticity, and cell-state transitions that are obscured in bulk analyses. Single-cell DNA sequencing (scDNA-seq) is the most direct method for revealing tumor clonal architectures at the cellular level, but its widespread adoption is constrained by technical bottlenecks, including ultralow DNA input per cell, amplification-induced artifacts, and high cost per cell [5]. To circumvent these issues, some suboptimal methods have been developed, such as isolating clonal cell populations from tissues or in vitro culture, followed by bulk whole-genome/exome sequencing (WGS/WES) [4,6,7]. However, these methods retain the fundamental flaws of bulk sequencing, thereby compromising their utility in deconvolving tumor heterogeneity. Single-cell-targeted sequencing offers another cost-effective approach, but its utility is limited by the number of detected cells and genetic loci [8,9], as well as dependency on bulk sequencing-derived prior knowledge for panel design. In addition, the technologies mentioned above lack the capacity for simultaneous interrogation of cancer cell genotypes and phenotypes [10].

Although techniques for sequencing both DNA and RNA from the exact same cell have been developed, this strategy remains technically challenging in human samples [11–14]. Some methods have been devised to infer the relationship between scDNA-seq and single-cell RNA sequencing (scRNA-seq), but still on different cells [15,16]. One low-resolution method for linking cell genotype and phenotype involves predicting DNA copy number variations (CNVs) based on scRNA-seq data, followed by hierarchical clustering of the CNVs to infer clonal evolution [17–19]. However, this method implicitly assumes a linear evolutionary model, where branch points in the resulting dendrogram are interpreted as chronological events, thereby overlooking the confounding effects of convergent evolution, copy number neutral losses of heterozygosity (LOH), and selective pressures that may distort the relationship between CNV distance and temporal acquisition. Although the

distance dendrograms in such analyses resemble phylogenetic trees, they actually fail to reflect clonal evolution [20].

It is technically feasible to directly detect somatic mutations in RNA sequencing reads from scRNA-seq data, which enables cell lineage tracing through transcriptional characteristics with annotation information. However, mutation detection and clonal inference from scRNA-seq data are limited by many factors such as differential expression, allelic imbalance, RNA editing, limited sequencing coverage or depth, sequencing artifacts, etc. [21]. It should be noted that, although RNA editing has been reported as an important phenotypic feature of cells that may be used for subclone identification [22,23], large numbers of RNA editing events may mask sparse genomic evidence in scRNA-seq data.

To address these challenges, we developed a full-process computational toolkit, scClone, for mutation detection and clonal evolution analysis based on single-cell transcriptome data. scClone will directly process raw sequencing reads to detect somatic mutations, impute drop-outs, and visualize clonal structures and evolutionary relationships. Our toolkit is not limited by methods of RNA library preparation or sequencing platforms, and it effectively takes advantage of single-cell transcriptomic annotations and bulk sequencing-derived mutational signatures. Currently, single-cell transcriptome analysis has matured significantly, supported by a vast volume of data, and now resembles a sprawling tree with intricate branches and lush foliage (Figure 1A). Clonal evolution, a cornerstone of cancer research, mirrors the intricate and interwoven root system of a thriving tree, where hidden branches fuel the tree's resilience and expansion. Our approach will add a new dimension for deconvolution of clonal architecture with single-cell transcriptomic profiling, provide new insights into tumor evolution, and hopefully contribute to future clinical applications.

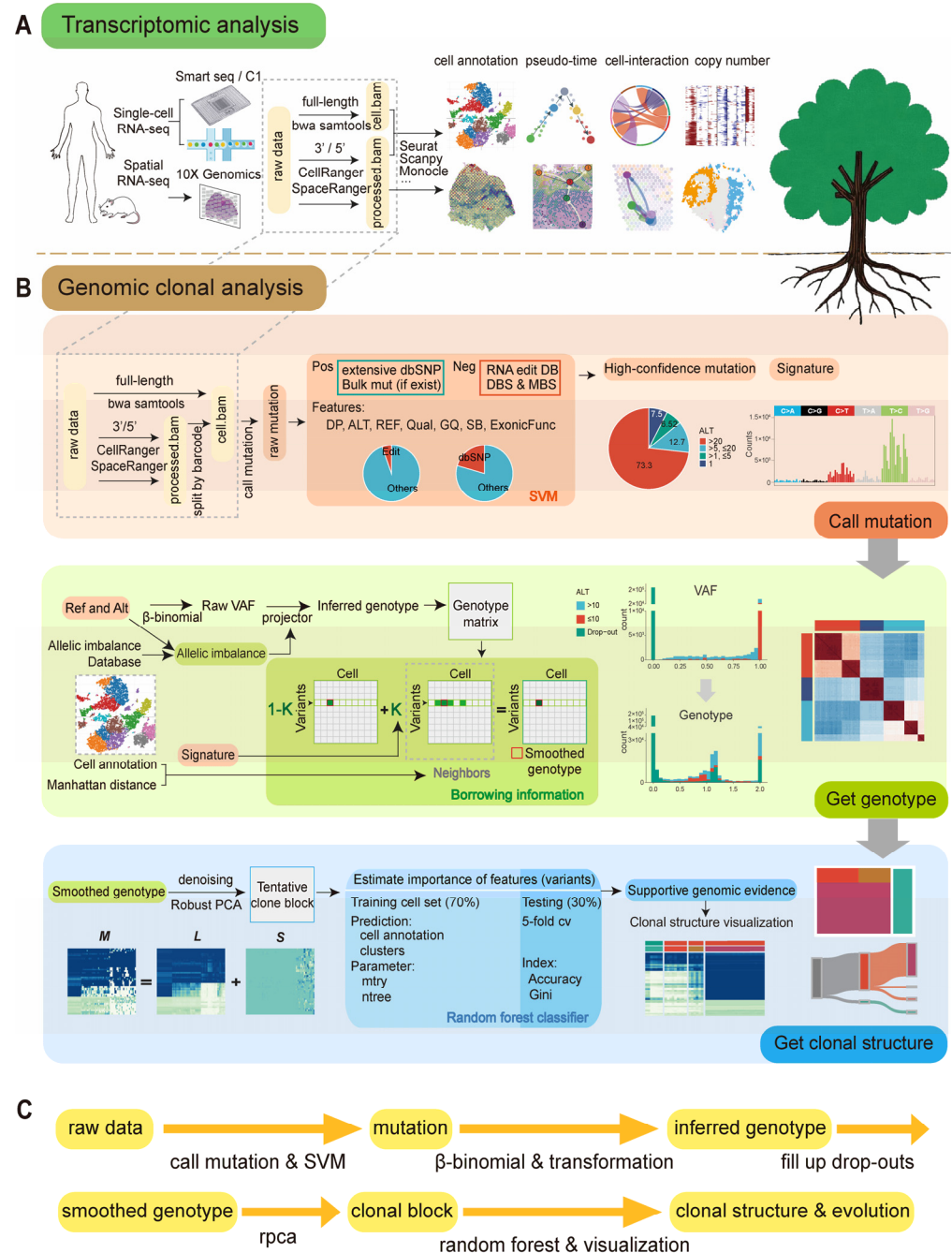


Figure 1. Workflow of scClone and its relationship with single-cell transcriptome analysis. (A) Schematic diagram of single-cell and spatial transcriptome sequencing processes and bioinformatics analysis methods. (B) Detailed workflow of scClone. SVM (support vector machine); DBS (doublet base substitutions); MBS (multi-base substitution); VAF (variant allele frequency); Qual (phred-scaled quality score); GQ (genotype quality); GQX (empirically calibrated genotype quality score for variant sites); ExonicFunc (exonic function); *M*, *L*, and *S* represent the optimized matrix, the low-rank matrix, and the sparse matrix, respectively. (C) Simplified workflow of scClone.

2. Results

2.1. Benchmark: Assessing the Effectiveness and Usability of scClone

To assess the efficiency of our toolkit, we initially used single-cell transcriptomic data of myeloma cells obtained via C1 platform as a benchmark [24]. The dataset included eight samples from six patients, including bone marrow, additional extramedullary (EM) ascites and pleural effusion samples. All cells were sorted using CD138-based flow cytometry

and data were derived from the same batch. Here we discarded cell annotations and processed all myeloma cells in parallel to infer the clonal structure. Given that these myeloma cells experienced different selective pressures across patients and tissues, eight samples served as a good reference for the clonal structure inference to assess our toolkit. We also selected another 10× Genomics dataset from a cutaneous squamous cell carcinoma (cSCC) patient, and its availability of bulk whole-exome sequencing (WES) data from paired tumor and adjacent normal tissues in the same patient provided further information for method assessment [25].

Regardless of the scRNA-seq library preparation chemistries, the majority of the raw mutations were located in intronic, intergenic and UTR3 regions, followed by exonic regions (Figure 2A and Figure S1A). This mutation distribution pattern was generally consistent with traditional genomic mutational profiling, with the larger genomic fractions of intronic and intergenic calls arising from incompletely spliced pre-mRNA, single-cell 3'/5'UTR capture and amplification artifacts. Additionally, both datasets exhibited a high prevalence of T > C and C > T mutations (Figure 2B and Figure S1B). Although previous studies have shown that clock-like and DNA damage signatures can also lead to C > T and T > C mutations, the substantial number of mutations observed here suggests that a considerable proportion of these mutations are likely attributable to RNA-editing events [22,26,27]. The full-length transcriptome data from the C1 platform exhibited an average of ~50,000 mutations per cell (Figure 2C), whereas the 10 × 3' sequencing data showed only ~3000 mutations per cell (Figure S1C). Upon analyzing the cumulative mutation counts across all single cells, we observed a near-linear relationship between total mutation count and cell number, which suggested that mutations were rarely shared among cells (Figure 2C and Figure S1C).

After processing with SVM, the proportion of mutation sites with read depth > 20 increased from 29.5% to 73.3%, whereas the proportion of mutations with a depth of 1 decreased from 48.3% to 7.5% (Figure 2D). These retained low-depth mutations might subsequently be utilized to infer the genotypes of drop-outs. The filtering also led to a notable reduction in T > C and C > T substitutions, which were primarily attributed to RNA editing (Figure 2E). As the cSCC samples had available WES data, it was observed that the mutational signature obtained after the SVM filtering more closely resembled the signatures derived from the bulk data, with cosine similarity raised from 0.47 to 0.79, which indirectly reflected the reliability of scClone in calling mutations and the usability of transcriptional mutation signatures (Figure 2F,G). To assess the usability of scClone in mutation detection, we artificially introduced 1000 mutations into randomly selected myeloma cells and performed six independent replicates (Supplementary Materials). Across all categories of mutations, scClone consistently achieved reproducibility rates exceeding 85% (Figure 2H). For the cSCC, we evaluated mutation reproducibility directly against bulk WES data; among the 880 mutations, 78 (8.9%) were detected in the scRNA-seq data processed by scClone (Figure 2I,J). Given the extremely low coverage of the 3'/5' sequencing regions, this level of concordance confirmed the adequacy of scClone for mutation detection. However, the limited mutations of cSCC epithelial cells in 3' sequencing data precluded further analysis. We further evaluated the latest scRNA-seq somatic mutation detection tool SComatic in the cSCC dataset and all mutations from SComatic were retained within the scClone mutation list. However, the limited yield rendered these calls insufficient for robust downstream clonal-evolution inference (Figure S2A–H).

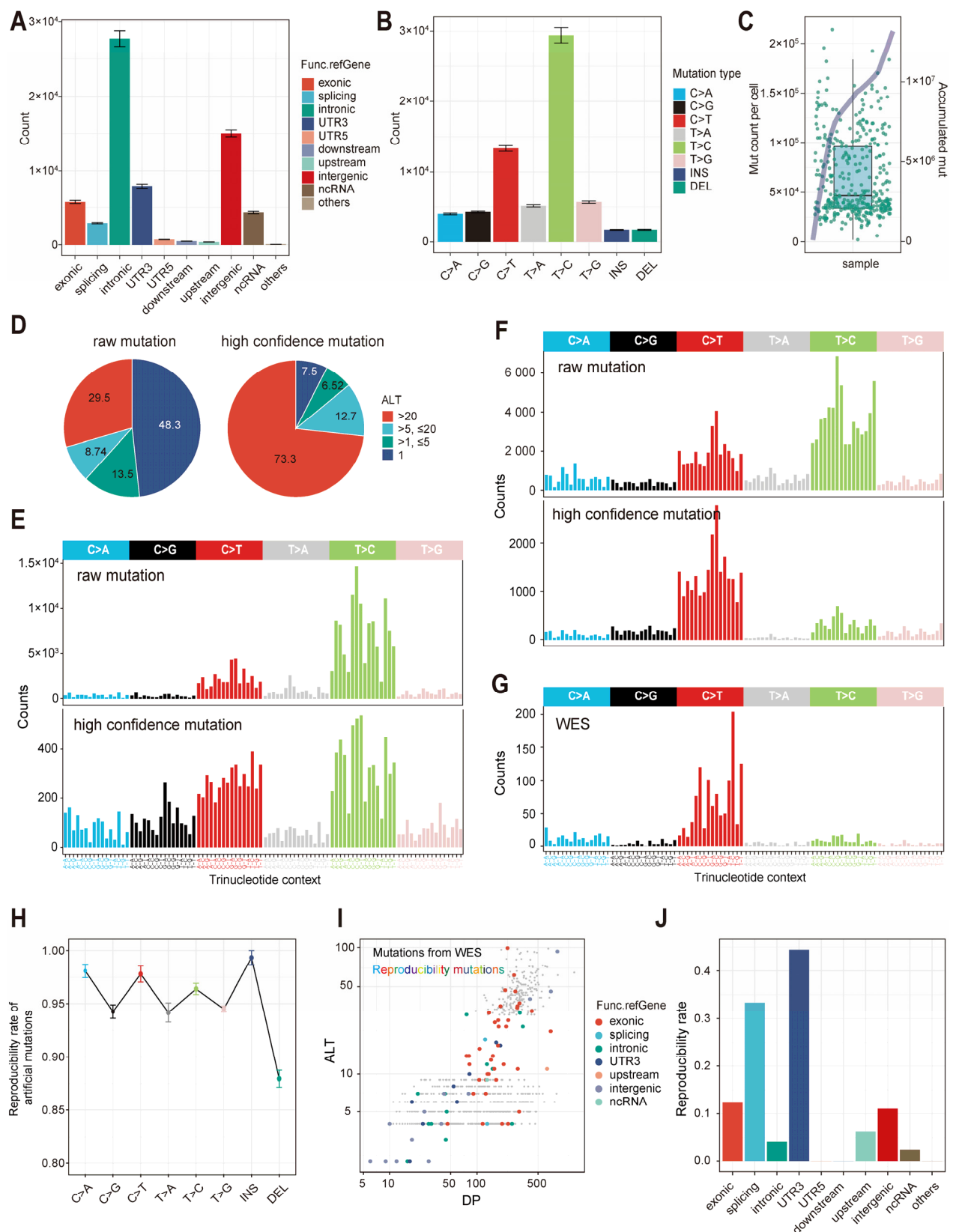


Figure 2. Detection of mutations in the transcriptome by scClone. (A) Statistics of functional annotation for raw mutations in myeloma cells detected by scClone. Error bars represent standard errors. (B) Statistics of the nucleotide substitution types for raw mutations in myeloma cells. Error bars represent standard errors. (C) Number of mutations in myeloma cells. Box and scatter plots show the number of mutations in each myeloma single cell (y-axis on the left), and the line shows the cumulative number of mutations in all myeloma cells (y-axis on the right). (D) Fractions of mutations

with ALT reads before and after SVM filtering in myeloma dataset, with raw mutations on the left and high-confidence mutations on the right. (E) Mutational signatures in 96 trinucleotide contexts before and after SVM filtering in myeloma cells. (F) Mutational signatures in the cutaneous squamous cell carcinoma (cSCC) dataset before and after SVM filtering. (G) Mutational signatures in cSCC via bulk-level exome sequencing. (H) Reproducibility of artificially introduced mutations in myeloma cells. The standard deviation from an independent repeated experiment has been labeled. (I) Mutations identified by WES in cSCC. Colored dots indicate mutations detected by scClone. (J) Reproducibility of scClone-detected mutations in cSCC relative to the WES.

In analysis of myeloma samples, the inferred genotypes exhibited obvious heterozygous peaks compared with raw VAF values (Figure 3A,B). The Manhattan distances of these inferred genotypes revealed indistinct clusters among cells from different sample origins (Figure 3D). Interestingly, after erasing cell annotation and filling in drop-outs based on the computed distances, the heatmap depicting the processed genotypes became more compact, with more pronounced differences among samples (Figure 3C–F). These outcomes could be attributed to the reliability of mutations, the rational inference of genotypes, and the effectiveness of the Manhattan distance in reflecting cellular clonal relationships. We assessed the clustering performance based on the progressively refined genotypes and found that the silhouette score improved in almost all samples (Figure 3G). Additionally, we evaluated the concordance between clusters and samples using various scores (CH, DB, AR, FM, and NMI), which indicated that the progressively refined processes increasingly approximated the true clone structure (Figure 3H). We also compared the impact of different thresholds for defining neighboring cells on clustering performance and found that scClone was not sensitive to these thresholds (Figure S3).

Next, we compared scClone with PhylinSic [20], another toolkit that also infers clonal evolution de novo from single-cell transcriptomes. In both datasets—pre- vs. post-treatment myeloma and primary vs. metastatic myeloma—scClone achieved overall superior clonal identification, accurately detecting clonal lesions from the same patient across different time and locations, and delivered better accuracy, recall, and F1-score ($p < 0.05$) (Figure 3I). We further assessed the clonal identities of single cells and determined the optimal hyperparameters to generate a low-complexity clonal structure within the random forest framework (Figure S4A–F). Five-fold cross-validation indicated the probability that cells from different samples were correctly classified (Figure S4D).

Ultimately, eight mutation-based tumor clonal clusters were formed (Figure 3J,K). Utilizing sample origins as independent clonal labels for orthogonal validation, we found that solely relying on scClone from raw data to final clonal identification without any sample information, six out of the eight genomic clonal clusters (cluster 2, 4, 5, 6, 7, and 8) accurately reflected their respective sample origins. Specifically, cells from different tissues of the same patient were clearly distinguished: the bone marrow (MM02) and pleural effusion samples (MM02EM) from patient MM02 were assigned to different clusters. The same situation was observed in the two samples from patient MM34. However, clonal clusters 1 and 3 were composed of cells from multiple sample origins. This phenomenon might necessitate the incorporation of additional genomic evidence or cell numbers to enhance the robustness of scClone and thereby improve the accuracy and resolution of clonal analysis.

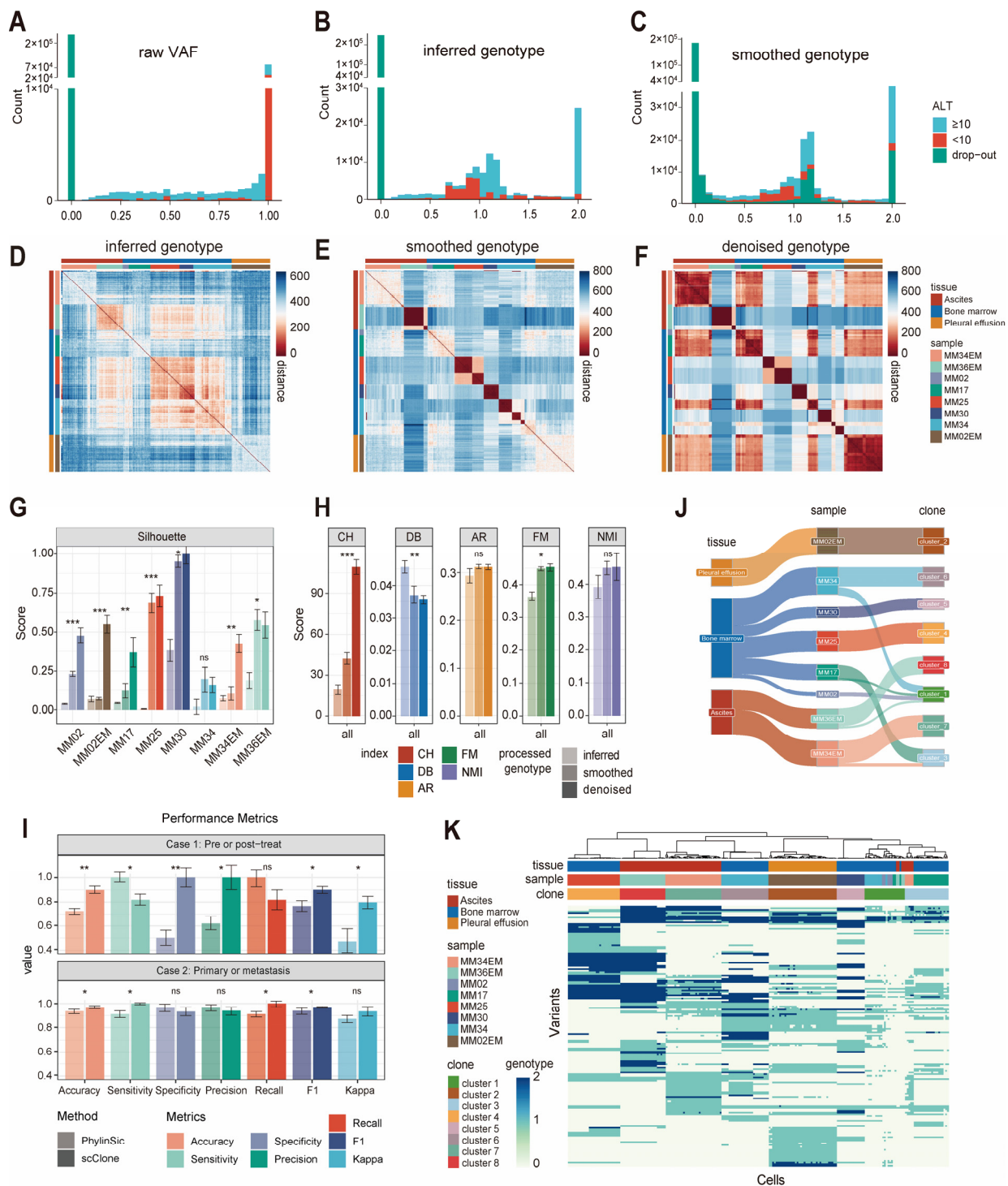


Figure 3. Evaluation of scClone performance in the myeloma dataset. **(A)** Raw VAF of high-confidence mutation sites. VAF ranges from 0 to 1. **(B)** Inferred genotypes. Genotype 0: wild-type; 1: heterozygous mutation; 2: homozygous mutation. **(C)** Smoothed genotypes after filling in drop-outs. **(D–F)** Heatmap of Manhattan distances between cell genotypes after different processing steps: inferred genotypes **(D)**, smoothed genotypes **(E)**, and denoised genotypes **(F)**. **(G)** Silhouette scores for different samples, with most samples showing improved scores after each of the three processing steps for genotypes. **(H)** Evaluation of genotype clustering similarity to sample clustering using multiple metrics during the stepwise analysis. AR: Adjusted Rand; CH: Calinski–Harabasz; DB: Davies–Bouldin; FM: Fowlkes–Mallows; NMI: Normalized Mutual Information. **(I)** Benchmarking scClone against Phylisic for clonal identification in pre- vs. post-treatment and primary vs. metastatic tumors. **(J)** Sankey diagram showing the relationship between tissue origin, sample identity and clonal clusters from scClone. **(K)** Visualization of scClone output, with each row representing a mutation

and each column representing a cell. Heatmap colors indicate genotypes. The standard deviation from an independent repeated experiment has been labeled. ANOVA (analysis of variance) test results are shown in (G,H). Student's *t*-test results are shown in (I) (ns: $p > 0.05$; *: $p \leq 0.05$; **: $p \leq 0.01$; ***: $p \leq 0.001$).

2.2. Identification of Subclones Among Single Cells by scClone

Aiming to assess the ability of scClone in detecting clonal structures in tumor cells, we applied scClone to full-length single-cell transcriptomic data of 800 cells from a hepatocellular carcinoma patient (HCC patient 2) [8]. According to the heatmap generated by scClone, three genomic mutational clusters were identified within the tumor cell population, clearly showing a stepwise accumulation of mutations (Figure 4A,B). Tumor cell cluster 1 tended to represent an early subclone with the lowest degree of genetic variation among tumor cells, and exhibited genotypes more closely resembling those of immune cells such as macrophages and T cells. In contrast, tumor cell cluster 3 exhibited the highest mutation count and represented the most recently evolved subclone, which may have acquired a selective advantage and subsequently expanded. Tumor cell cluster 2 occupied an intermediate stage in the evolutionary trajectory from tumor cell cluster 1 to tumor cell cluster 3. By examining the mutated genes, we found that tumor cells progressively acquired non-synonymous mutations in genes such as KTN1, ALB, APOB, CDK11A, and CDK11B (Figure 4C). We then analyzed single-cell target sequencing (scTarget-seq) data from the HCC patient 2 [8]: 50 cells and 55 variant positions yielded a cell-by-mutation matrix that resolved two tumor subclones whose composition resembled the tumor cell clusters 1 and 3 identified by scClone (Figure S5). Given the limited throughput of scTarget-seq, the intermediate tumor cell 2 was not captured. Despite the scRNA-seq and scTarget-seq sampling different tumor sectors, the convergent clonal picture corroborates the robustness of the scClone reconstruction.

We then mapped the clonal information derived from scClone onto a two-dimensional tSNE map, where the distances between cells reflected transcriptomic similarities (Figure 4D). Notably, tumor cell clusters 1 and 3 were mapped to two distinct transcriptomic clusters on the tSNE map, while tumor cell cluster 2 demonstrated overlap with both clusters. Interestingly, tumor cell cluster 2 represented an intermediate state in both clonal evolution and transcriptomic profiles. This observation suggested a close relationship among the three clonal clusters with different genetic variations and the transcriptional expression profiles. Indeed, it was reasonable to observe distinct expression profiles among tumor subclones, given that subclone-specific mutations might exert varying functional impacts on the transcriptomic landscape.

We further investigated the relationship between scClone-derived clonal clusters and multiple concepts derived for single-cell transcriptomes. Mapping clonal cluster information onto a branched trajectory of pseudo-time analysis, we observed that tumor cell cluster 3 was concentrated on the left branches, while tumor cell cluster 1 and 2 did not show significant enrichment, suggesting a lack of consistency between mutation-based clonal clusters and pseudo-time trajectories (Figure 4E). One explanation was that pseudo-time analysis like Monocle2 identified continuous gene expression patterns to infer “time,” which might not be on the same timescale compared with mutation accumulation. Thus, the two approaches provided complementary information for better understanding of “time” and “evolution” at multiple levels. We further calculated CNVs from the gene expression profiles by inferCNV [28], and the results were partially aligned with scClone clonal clusters. Tumor cell cluster 3 exhibited widespread chromosome 19 deletions and frequent amplifications on chromosomes 8 and 12, while no significant CNVs were detected in tumor cell cluster 1

and 2 (Figure 4F). As clonal clusters from scClone were based on genomic mutations, the resolution was relatively higher than CNV clusters from transcriptional profiles.

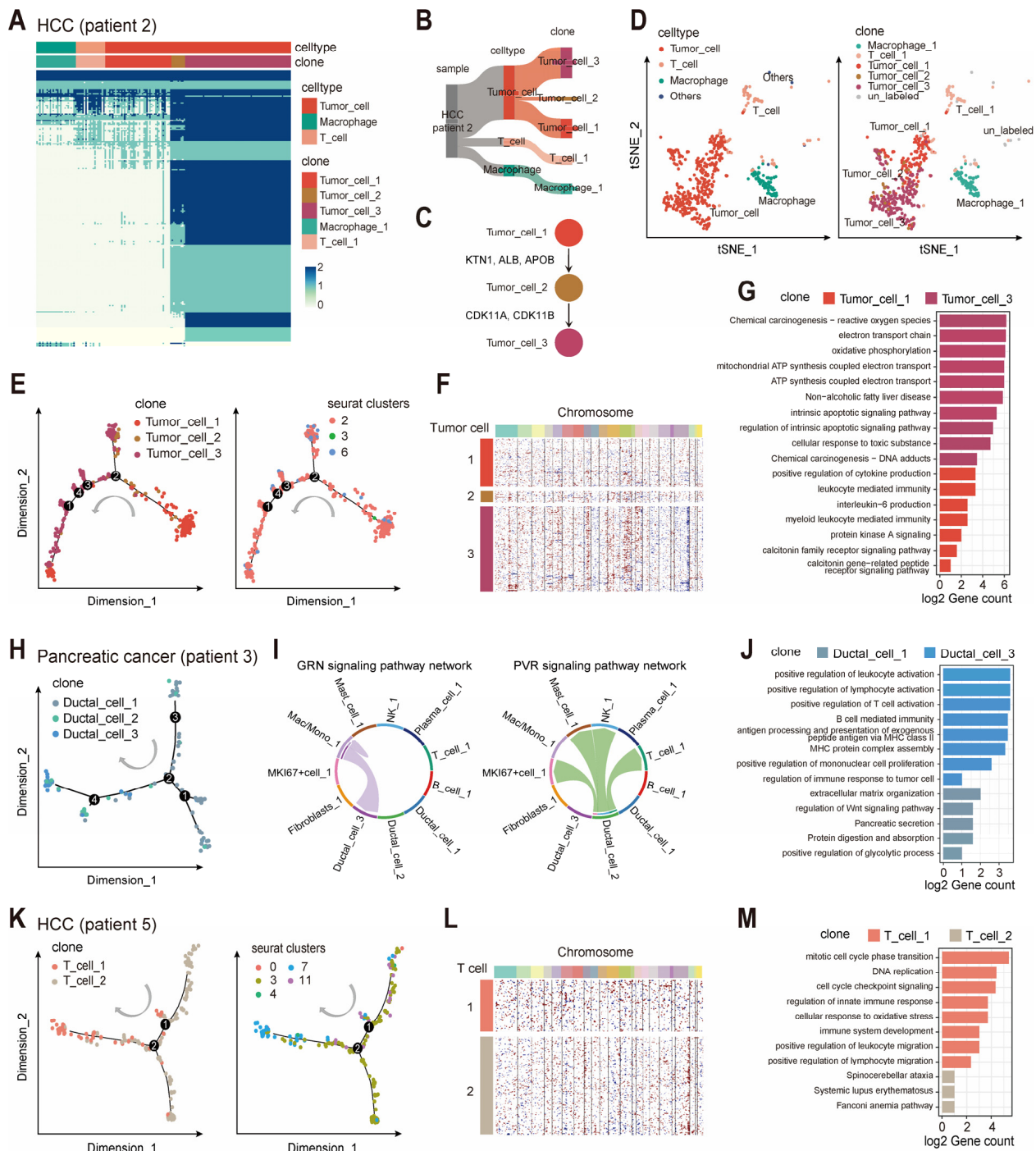


Figure 4. Clonal clusters identified via scClone have functional differences. (A) "Variant-cell" map of an HCC patient (patient 2) based on scClone. (B) Sankey diagram showing cell types and clonal clusters in HCC patient 2. (C) Schematic of tumor clonal progression in HCC patient 2 with non-synonymous key mutations. (D) tSNE map of HCC patient 2, with cell type (left) and projected clonal cluster information (right). (E) Pseudo-time trajectory of tumor cells in HCC patient 2, with clonal cluster (left) and transcriptional cluster information (right). (F) Whole-genome copy number variations (CNV) inferred from a single-cell transcriptome for tumor cells in HCC patient 2, with each row representing a cell and each column representing a chromosomal bin. (G) Enrichment analysis of differentially expressed genes between tumor clones in HCC patient 2. (H) Pseudo-time trajectory of ductal cells in a pancreatic cancer patient (patient 3), with clonal cluster information. (I) Interaction

of signaling pathways among different cell clones in pancreatic cancer patient 3. **(J)** Enrichment analysis of differentially expressed genes between clonal clusters in pancreatic cancer patient 3. **(K)** Pseudo-time trajectory of T cells in HCC patient (patient 5), with clonal cluster (left) and transcriptional cluster information (right). **(L)** Whole-genome CNV inferred from single-cell transcriptome for T cells in HCC patient 5. **(M)** Enrichment analysis of differentially expressed genes between clonal clusters of T cells in HCC patient 5.

From enrichment analysis of clonal cluster-specific expressed genes, we found functional differences in metabolism between tumor cell clusters 1 and 3. Tumor cell cluster 1 showed enrichment with immunity and cell signaling, involving leukocyte-mediated immunity and production of cytokines and interleukin-6 (Figure 4G). These immune-related signatures in cluster 1 may have important clinical implications—interleukin-6 is a well-established pro-inflammatory cytokine associated with aggressive tumor phenotypes and poor prognosis in hepatocellular carcinoma (HCC), suggesting that tumor cell cluster 1 with higher infiltration or activity may carry an increased risk of recurrence or distant metastasis [29,30]. In contrast, tumor cell cluster 3 focused more on energy metabolism, cellular stress and apoptosis, as well as several disease-related pathways like non-alcoholic fatty liver disease and chemical carcinogenesis. Tumor cell cluster 3 highlights a potential link to metabolic-associated HCC subtypes, which are increasingly prevalent clinically and require tailored therapeutic strategies; meanwhile, the emphasis on cellular stress and apoptosis may imply cluster 3 cells are more vulnerable to metabolic-targeted agents or oxidative stress-inducing therapies, providing a direction for precision treatment [31]. By investigating cell–cell interactions by CellChat [32], we found that tumor cell 3 had more MHC-I signaling interactions with tumor cell cluster 1 and macrophages, and tumor cell cluster 2 had PECAM1 signaling interactions with macrophages and exhibited a strong WNT signaling pathway interaction with other cell clusters (Figure S6A,B). Enhanced MHC-I-mediated crosstalk suggests immunotherapy responsiveness, as MHC-I supports antigen presentation and T cell-activation serves as a predictive biomarker for immune checkpoint inhibitor patient selection. Additionally, cluster 2's prominent WNT signaling implies a role in tumor progression and stemness, closely linked to HCC recurrence and poor survival [33]. Collectively, these cluster-specific features reveal HCC's molecular heterogeneity and offer actionable insights for clinical stratification, prognostic assessment, and personalized therapy.

We further applied scClone to pancreatic cancer scRNA-seq data from the 10× Genomics platform (patient 3), which contained more cells and retained greater transcriptomic diversity, including ductal cells, fibroblasts and multiple types of immune cells [34] (Figure S7A–D). However, due to the lower amount of mRNA captured per cell in the 10× 5' chemistry and the requirement of sufficient mutations for clonal identification, a higher proportion of single cells were filtered out by scClone compared to full-length data. Ductal cells were divided into three clonal clusters by scClone, with a gradual accumulation of mutations across these clusters (Figure S7A). This dynamic process was also evident in pseudo-time trajectory analysis, where ductal cell cluster 1 was enriched in the early stages and ductal cell cluster 3 at the terminal end (Figure 4H). We also uncovered different transcriptional activity levels of the GRN and PVR signaling pathways among the three clonal clusters (Figure 4I and Figure S7E). Ductal cell cluster 1 was primarily enriched with metabolic pathways, whereas ductal cell cluster 3 exhibited substantial activity related to immunity and antigen presentation, suggesting ductal cells derived from cluster 3 could be more responsive to immunotherapies, providing a potential marker for treatment selection (Figure 4J).

2.3. *scClone Describes the Evolutionary Trajectories of Immune Cells*

The concept of clones in immune cells, particularly T cells and B cells, is traditionally defined based on TCR and BCR clones, which are determined by the rearrangement of the VDJ regions in the genome. These clones are used to describe the evolution of T cells and B cells to recognize specific antigens. In fact, any proliferating cells, whether tumor cells or non-tumor cells, will inevitably accumulate mutations during the DNA replication process. In a full-length single-cell transcriptomic dataset from another HCC patient (patient 5), scClone identified two clusters of T cells (Figure S8A–D). These two T-cell clusters exhibited transcriptomic differences in pseudo-time analysis (Figure 4K), and their differences were also supported by copy number inference, with T-cell cluster 2 harboring genomic amplification at 19q regions (Figure 4L). For enrichment analysis of genes differentially expressed between the two clusters, T-cell cluster 1 was primarily involved in cell proliferation and immune activity, indicating its participation in tumor immune infiltration, while T-cell cluster 2 was associated with limited pathways (Figure 4M). Additionally, T-cell cluster 1 exhibited more interactions with other cells through the CD96 signaling pathway and communicated with tumor cells via EPGN (Figure S8E). Notably, CD96 signaling dysregulation in T cells has been linked to impaired anti-tumor immunity and poor clinical outcomes, implying that targeting this pathway could enhance the therapeutic effect for this patient, providing a personalized treatment direction [35,36].

2.4. *scClone Enables Clonal Cluster Identification in Spatial Transcriptomics*

The high similarity between single-cell and spatial transcriptomics data allows easy application of scClone to spatial transcriptomics data. The key difference between the two approaches lies in the resolution, as spatial transcriptomics are based on ROI (region of interest) spots, each containing multiple cells. This makes it difficult to precisely define the cell types within a spot. The requirement for enough mutations in a spot for clonal inference inevitably leads to the filtering out of some spots, resulting in the loss of information at these sites, which cannot be recovered. These two factors make clonal inference in spatial transcriptomics more challenging and lower in resolution.

We first demonstrated the ability of scClone to identify clonal clusters in the ovarian cancer spatial atlas. For two cases of ovarian cancer spatial transcriptomics data from patients 3 and 8 [37], scClone identified two subclones in each patient that showed distinct clustering on pathological sections (Figure 5A,G). In patient 3, two clonal clusters were consistent with the histological sections, suggesting a difference in tumor morphology for the two subclones (Figure 5B). UMAP results of the expression profiles revealed the differential expression between the two subclones in patient 3 (Figure 5C), together with significantly different enrichment along the pseudo-time trajectory (Figure 5D). Cluster 1 tended to exhibit higher tumor purity and malignancy as evidenced by more amplifications of chromosome 6 in copy number inference, while cluster 2 appeared relatively normal (Figure 5E). Cluster 1 involved the activation of immune cells, antigen processing, apoptosis and cell cycle regulation, while cluster 2 was associated with metabolism and homeostasis (Figure 5F). This evidence indicated that tumor clone regions derived from cluster 1 have faster tumor progression and a higher antigen response.

In ovarian cancer patient 8, we similarly identified that cluster 1 had more accumulated mutations and a relatively normal cluster 2 (Figure 5G). Cluster 1 was prominently clustered in the upper region of the histological section, while cluster 2 spanned across a much larger region (Figure 5H). Both clonal clusters showed significantly different expression and pseudo-time enrichment (Figure 5I,J). In addition to differences in mutations revealed by scClone, these two subclones displayed clear distinctions in copy number inference, with cluster 1 exhibiting amplifications of chromosomes 3, 8, 13, 20 and 21, along with deletions

of chromosomes 10 and 22 (Figure 5K). Cluster 1 was characterized by RNA transport and modification, nucleotide metabolism and intercellular signaling pathways such as the Wnt signaling pathway. Cluster 2 was mainly enriched with antigen processing and presentation, and immune activation (Figure 5L). Thus, the clonal structure identified in spatial transcriptomics by scClone was generally consistent with the histological sections and provided special biological insight.

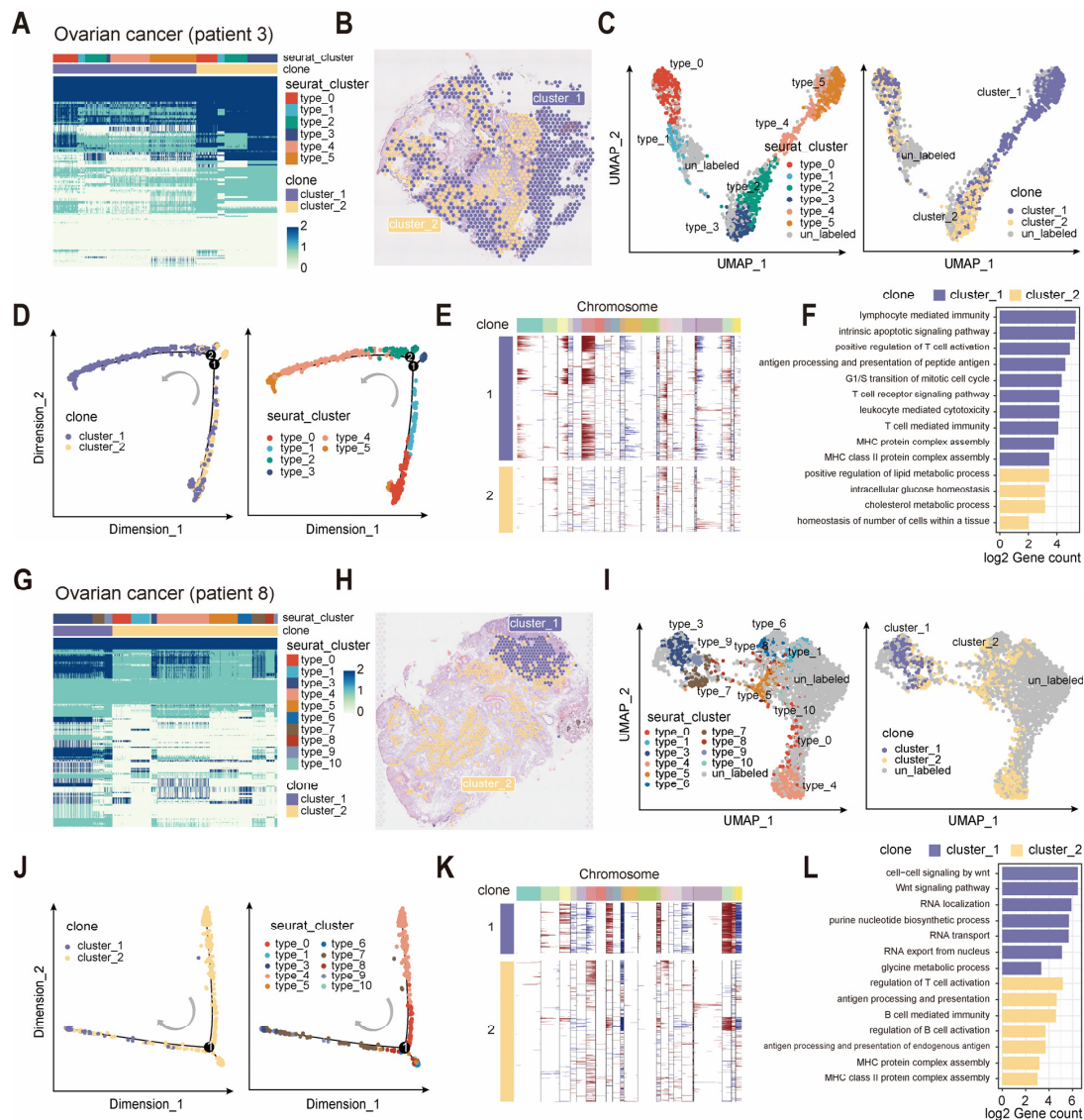


Figure 5. scClone enables clonal cluster identification in spatial transcriptomics. (A) "Variant-cell" map of ovarian cancer patient (patient 3) based on spatial transcriptome from scClone. (B) Clonal structure of ROI spots displayed on pathological sections from patient 3. (C) UMAP of single cells from patient 3, with cell type (left) and clonal cluster information (right). (D) Pseudo-time trajectory of ROI spots in patient 3, with clonal cluster (left) and transcriptional cluster information (right). (E) Whole-genome CNV inferred from single-cell transcriptome for patient 3, with each row representing a spot and each column representing a chromosomal bin. (F) Enrichment analysis of differentially expressed genes between clonal clusters in patient 3. (G) "Variant-cell" map of ovarian cancer patient (patient 8). (H) Clonal structure of ROI spots displayed on pathological sections from patient 8. (I) UMAP of single cells from patient 8, with cell type (left) and clonal cluster information (right). (J) Pseudo-time trajectory of ROI spots in patient 8, with clonal cluster (left) and transcriptional cluster information (right). (K) Whole-genome CNV inferred from single-cell transcriptome for patient 8. (L) Enrichment analysis of differentially expressed genes between clonal clusters in patient 8.

2.5. Integration of scClone and Transcriptomic Information Reveals High-Resolution Clonal Structures

The results from the two spatial cases above demonstrated a degree of consistency between the genotype clusters inferred by scClone and the clusters of spatial spot transcriptomes (Figure 5C,I). However, unlike scRNA sequencing, the low utilization rate of ROI spots in spatial transcriptomics for scClone limited the resolution of clonal structures, as shown in a cutaneous squamous cell carcinoma sample (cSCC patient 6), where the sparse spots revealed two major clonal structures (Figure S9A–C). Consistent with the previous process, scClone identified two clonal clusters in another sample of cSCC (cSCC patient 4) (Figure S9D–F) [25]. Fortunately, scClone retained a high proportion of ROI spots here, and the sufficient spots allowed us to re-infer the clonal structure by utilizing transcriptomic annotations, ultimately obtaining nine subclonal clusters among seven transcriptomic groups (Figure 6A). Type_1_1 and type_1_2, which belonged to expressional cluster type_1, were identified as two distinct subclones by scClone based on their genotype differences. These two tumor subclones demonstrated consistency with their spatial positions, which cannot be distinguished by either gene expression or transcriptome-inferred CNVs (Figure S9G,H). Further enrichment and cell interaction analysis revealed that these two tumor subclones shared similar activities in the ACTIVIN and IL1 signaling pathways, while type_1_2 showed higher IL-10 and calcitriol activity with other spots (Figure 6B and Figure S9I). Type_1_1 exhibited high activity of pathways related to skin development and keratinocyte function, indicating a closer resemblance to normal skin cell functions. In contrast, type_1_2 was characterized by lipid metabolism, the NF- κ B immune signaling pathway and immune cell activation (Figure 6C). Here we demonstrated that the integration of scClone and spatial transcriptomic information could provide a more in-depth understanding of the genetic and transcriptomic features of tumor subclones, thereby facilitating the diagnosis of complex clonal architectures and highlighting the clinical potential of scClone in tumor precision medicine.

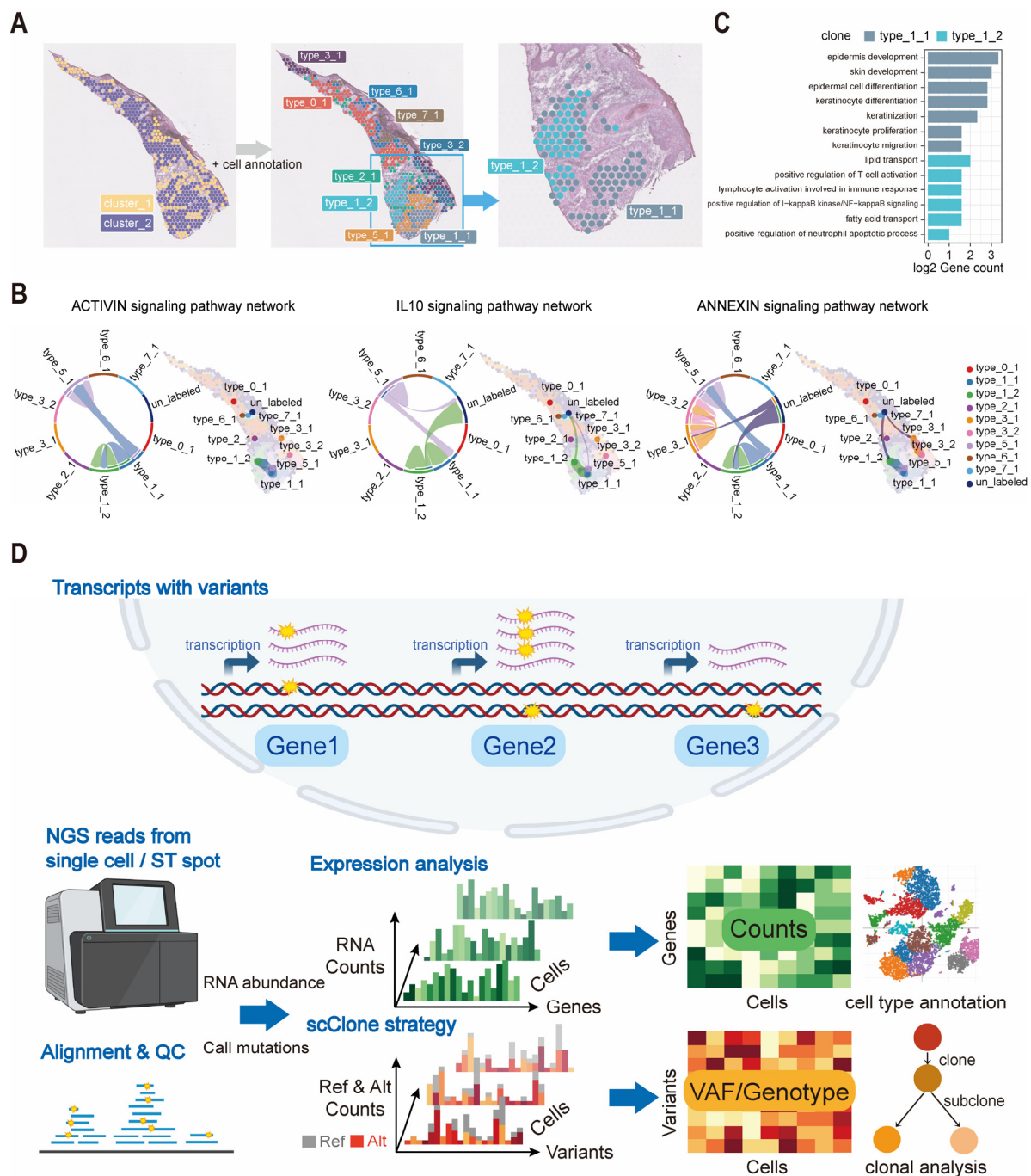


Figure 6. Functional interpretation of scClone-derived clonal clusters in spatial transcriptomics. (A) Spatial clonal structure of cSCC (patient 4) (left), clonal structure combined with transcriptomic clusters (middle), and zoomed-in view of the two clonal clusters of type_1 (right). (B) Interaction of signaling pathways between spatial transcriptomic spot clones in patient 4. (C) Enrichment analysis of differentially expressed genes between clone populations in patient 4. (D) The schematic diagram of the scClone strategy and its differences with the transcriptome expressional analysis workflow.

3. Discussion

Single-cell mutational profiling emerges as a promising approach to studying clonal evolution compared with bulk VAF-based methods, and several lineage reconstruction algorithms have been developed, such as OncoNEM [38], SCITE [39], SiFit [40], and SiCloneFit [41]. These tools are developed for single-cell genome or exome sequencing data instead of scRNA-seq data, and they are mainly designed for scenarios with a small number of tumor cells. There are also some bioinformatic solutions for mutational inference from

scRNA-seq, such as SCmut [42] and Cardelino [43], but they often require prior knowledge from matched bulk DNA sequencing data [43–46]. These strategies have not been widely used, as sampling bias between genomic and transcriptomic data may affect the reliability of clonal inference and assignment. More recently, some algorithms such as Monopogen [47] and SComatic [48] have been created for somatic mutation detection in scRNA-seq data, but they stop at variant detection: the sparse mutational landscapes they yield lack the statistical robustness for reliable reconstruction of clonal architecture and downstream evolutionary inference.

Our scClone is a completely de novo clonal structure inference toolkit for single-cell transcriptomics data, which is specially designed to avoid interference from RNA editing and allelic imbalance. It has a comprehensive set of functions, including mutation detection, clonal cluster identification, and clonal evolutionary analysis (Figure 6D). scClone could efficiently and simultaneously process thousands of single cells with different cell type identities, leveraging rich public scRNA-seq resources. Furthermore, it could be applied to spatial transcriptomics data with no need for prior knowledge from bulk or single-cell mutational profiling. scClone also incorporates the concept of mutational signatures from bulk sequencing and cell annotation from transcriptomics to intuitively present the clonal structure of a sample to users.

During the development of scClone, we have noticed that the number of raw mutations detected in scRNA-seq was often 4 to 5 orders of magnitude higher than mutations from bulk DNA sequencing, including a large number of sequencing errors and RNA-editing events. Even after strict quality control and filtering steps succeeded in preserving and revealing genuine mutations, the overlap between mutations identified by bulk sequencing and those derived from single-cell transcriptomes remained limited. This situation might arise from the fact that mutations occur constantly and randomly during cell division and most single-cell-specific mutations are too sparse to be detected in bulk sequencing with limited coverage. As evidence, the filtered mutations in our pipeline contain massive low-frequency mutations, which may be assigned to the neutral tail. Conversely, mutations detected by bulk genomic sequencing are rarely observed in single-cell transcriptomes due to various factors such as limited sequencing regions and coverage, differential expression and allelic imbalance. The reliable mutation detection from scRNA-seq remains a significant challenge today.

Designed to analyze mainstream single-cell and spatial transcriptomics datasets, the scClone workflow is compatible with prevailing single-cell transcriptomic data formats, features excellent extensibility, and can be directly applied to both mouse and human samples, aiming to ensure easy accessibility for various users. As a full-process computational tool for single-cell mutation detection, scClone consumes relatively substantial resources on personal servers; the resource consumption of all its processes in a cSCC sample (6000 cells) is provided in Figure S10A,B. Users can perform pre-screening to improve computational efficiency, such as filtering low-quality reads and excluding cells with low expression or unassigned cell types. Moreover, the loss of spatial resolution resulting from spot filtering remains an inherent limitation that the genome-based scClone framework is currently unable to resolve. One of the advantages is that spatial transcriptomics achieves higher mutation detection depth and higher mutation abundance within spots at the cost of cellular resolution (Figure S11A–D). Overall, our aim is to enable the exploitation of the genomic information hidden in the widespread scRNA-seq data, enabling researchers to link genetic and transcriptomic features and gain new insights into tumor initiation and progression.

4. Materials and Methods

4.1. Data Acquisition

No new sequencing data are generated in this study and all data are obtained from the following sources: The C1 single-cell transcriptome data and single-cell target sequencing data for hepatocellular carcinoma (HCC) were obtained from our previous work, easy to access with Gene Expression Omnibus (GEO) and Sequence Read Archive (SRA) under accession numbers GSE146115 and PRJNA606993 [8]. Myeloma data for the benchmark were obtained from GEO under accession numbers GSE106218 and GSE110499 [24]. The single-cell expression data of pancreatic cancer were acquired from GEO under accession GSE197177. Raw data of pancreatic cancer were obtained from Genome Sequence Archive under accession numbers HRA004556 (scRNA-seq) and HRA004625 (WES) [34]. The single-cell and spatial transcriptome data of cutaneous squamous cell carcinoma (cSCC) were acquired from GSE144240 [25]. The spatial transcriptome data of ovarian cancer were obtained from GSE211956 [37].

4.2. scClone Workflow

The first step of our tool is the detection and filtering of mutations (Figure 1B,C). We develop distinct pipelines for two major types of scRNA-seq raw data, full-length and 3'/5' RNA sequencing, to obtain mutations per cell. These mutation sites are filtered using a support vector machine (SVM) to obtain high-confidence mutations. In this process, mutations cataloged in the dbSNP database with high prevalence are used as the positive dataset, while mutations recorded in the RNA editing database without aa-changed [26,49] and multi-base mutations serve as the negative dataset for training the identification of unknown mutations. The training features are derived from the VCF files of raw mutations.

The second step of scClone is inference of the genotype of each cell at each genomic site based on the mutations (Figure 1B,C). Given the low sequencing depth, the VAF values for the mutations will be transformed into positive real numbers of 0, 1, or 2 through beta-binomial distribution and allelic imbalance-based transformation. For the estimation of allelic imbalance, we consider expression rate and the database of human allelic expression [50]. Due to differential expression and limited capture counts for RNA transcripts, a sparse mutation-by-cell matrix is formed for each sample. We also employ a method of borrowing information from neighboring cells to fill in the drop-outs. As it is generally believed that genomic clonal identity should not contradict cell type, the parameters for filling in the drop-outs and identifying neighboring cells take the transcriptomic cell annotation and mutational signatures into account. This process forms a more detailed and informative mutation-by-cell matrix.

The third step aims to further denoise the matrix and uncover the clonal structure derived from single-cell transcriptomics (Figure 1B,C). Exclusive genotypes detected in a very few cells will be filtered out as noise through a robust principal component analysis (RPCA) process to make the clonal structure more observable. Finally, the accuracy and Gini score of all mutations for the initial hierarchical clusters are assessed using a random forest method, and mutations consistent with the clusters and cell annotations are used to draw the clonal structure diagram and reconstruct evolutionary trajectories (see Supplementary Materials for details).

5. Conclusions

- This work introduces scClone, a full-process clonal evolution inference toolkit based on mutation detection that relies solely on single-cell transcriptome sequencing. It includes a reliable mutation detection pipeline, a series of genotype inference algorithms, and clonal structure visualization.

- scClone achieves promising results across various cell types from different platforms and is compared with mainstream transcriptome analysis methods.
- scClone can be applied to spatial transcriptomics and identifies subclonal structures on histological sections that traditional methods fail to detect.

Supplementary Materials: The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/ijms262311428/s1>. References [26,28,32,49,51–58] are cited in the Supplementary Materials file.

Author Contributions: S.B. (Investigation, Methodology, Analysis, Writing), X.S. (Supervision, Analysis, Editing), Z.C. (Investigation, Analysis), Z.-G.H. (Supervision, Resources, Editing). All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (82272969, 82472880, 82441040 and 82073116 to Z.-G.H.), the National Key Research and Development Program of China (2022YFA1302700 to Z.-G.H.), Natural Science Foundation of Shanghai (21JC1403200 to Z.-G.H., 25ZR1401192 to X.S.), 111 Project (B17029) and the Fundamental Research Funds for the Central Universities.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: This article does not generate any new sequencing data. The data presented in this study are available in NCBI GEO repository at <https://www.ncbi.nlm.nih.gov/geo/> (accessed on 1 November 2023), reference numbers GSE146115, GSE106218, GSE110499, GSE197177, GSE144240 and GSE211956, for more details in Materials and methods. The source code in this article is available at <https://github.com/monkeyBai96/scClone> (accessed on 1 July 2025), with R package (version 1.0) in <https://github.com/monkeyBai96/scClone/tree/main/package> (accessed on 1 July 2025).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Aparicio, S.; Caldas, C. The implications of clonal genome evolution for cancer medicine. *N. Engl. J. Med.* **2013**, *368*, 842–851. [CrossRef]
2. Venkatesan, S.; Swanton, C. Tumor Evolutionary Principles: How Intratumor Heterogeneity Influences Cancer Treatment and Outcome. *Am. Soc. Clin. Oncol. Educ. Book* **2016**, *35*, e141–e149. [CrossRef]
3. Frieda, K.L.; Linton, J.M.; Hormoz, S.; Choi, J.; Chow, K.K.; Singer, Z.S.; Budde, M.W.; Elowitz, M.B.; Cai, L. Synthetic recording and in situ readout of lineage information in single cells. *Nature* **2017**, *541*, 107–111. [CrossRef] [PubMed]
4. Su, X.; Bai, S.; Xie, G.; Shi, Y.; Zhao, L.; Yang, G.; Tian, F.; He, K.Y.; Wang, L.; Li, X.; et al. Accurate tumor clonal structures require single-cell analysis. *Ann. N. Y. Acad. Sci.* **2022**, *1517*, 213–224. [CrossRef]
5. Gawad, C.; Koh, W.; Quake, S.R. Single-cell genome sequencing: Current state of the science. *Nat. Rev. Genet.* **2016**, *17*, 175–188. [CrossRef]
6. Moore, L.; Cagan, A.; Coorens, T.H.H.; Neville, M.D.C.; Sanghvi, R.; Sanders, M.A.; Oliver, T.R.W.; Leongamornlert, D.; Ellis, P.; Noorani, A.; et al. The mutational landscape of human somatic and germline cells. *Nature* **2021**, *597*, 381–386. [CrossRef] [PubMed]
7. Lee-Six, H.; Obro, N.F.; Shepherd, M.S.; Grossmann, S.; Dawson, K.; Belmonte, M.; Osborne, R.J.; Huntly, B.J.P.; Martincorena, I.; Anderson, E.; et al. Population dynamics of normal human blood inferred from somatic mutations. *Nature* **2018**, *561*, 473–478. [CrossRef]
8. Su, X.; Zhao, L.; Shi, Y.; Zhang, R.; Long, Q.; Bai, S.; Luo, Q.; Lin, Y.; Zou, X.; Ghazanfar, S.; et al. Clonal evolution in liver cancer at single-cell and single-variant resolution. *J. Hematol. Oncol.* **2021**, *14*, 22. [CrossRef]
9. Tang, J.; Tu, K.; Lu, K.; Zhang, J.; Luo, K.; Jin, H.; Wang, L.; Yang, L.; Xiao, W.; Zhang, Q.; et al. Single-cell exome sequencing reveals multiple subclones in metastatic colorectal carcinoma. *Genome Med.* **2021**, *13*, 148. [CrossRef] [PubMed]
10. Lahnemann, D.; Koster, J.; Szczurek, E.; McCarthy, D.J.; Hicks, S.C.; Robinson, M.D.; Vallejos, C.A.; Campbell, K.R.; Beerenwinkel, N.; Mahfouz, A.; et al. Eleven grand challenges in single-cell data science. *Genome Biol.* **2020**, *21*, 31. [CrossRef]

11. Macaulay, I.C.; Haerty, W.; Kumar, P.; Li, Y.I.; Hu, T.X.; Teng, M.J.; Goolam, M.; Saurat, N.; Coupland, P.; Shirley, L.M.; et al. G&T-seq: Parallel sequencing of single-cell genomes and transcriptomes. *Nat. Methods* **2015**, *12*, 519–522. [[CrossRef](#)] [[PubMed](#)]
12. Dey, S.S.; Kester, L.; Spanjaard, B.; Bienko, M.; van Oudenaarden, A. Integrated genome and transcriptome sequencing of the same cell. *Nat. Biotechnol.* **2015**, *33*, 285–289. [[CrossRef](#)]
13. Nam, A.S.; Kim, K.T.; Chaligne, R.; Izzo, F.; Ang, C.; Taylor, J.; Myers, R.M.; Abu-Zeinah, G.; Brand, R.; Omans, N.D.; et al. Somatic mutations and cell identity linked by Genotyping of Transcriptomes. *Nature* **2019**, *571*, 355–360. [[CrossRef](#)]
14. Reuter, J.A.; Spacek, D.V.; Pai, R.K.; Snyder, M.P. Simul-seq: Combined DNA and RNA sequencing for whole-genome and transcriptome profiling. *Nat. Methods* **2016**, *13*, 953–958. [[CrossRef](#)] [[PubMed](#)]
15. Li, R.; Ferdinand, J.R.; Loudon, K.W.; Bowyer, G.S.; Laidlaw, S.; Muyas, F.; Mamanova, L.; Neves, J.B.; Bolt, L.; Fasouli, E.S.; et al. Mapping single-cell transcriptomes in the intra-tumoral and associated territories of kidney cancer. *Cancer Cell* **2022**, *40*, 1583–1599.e10. [[CrossRef](#)] [[PubMed](#)]
16. Campbell, K.R.; Steif, A.; Laks, E.; Zahn, H.; Lai, D.; McPherson, A.; Farahani, H.; Kabeer, F.; O’Flanagan, C.; Biele, J.; et al. clonealign: Statistical integration of independent single-cell RNA and DNA sequencing data from human cancers. *Genome Biol.* **2019**, *20*, 54. [[CrossRef](#)]
17. Tirosh, I.; Izar, B.; Prakadan, S.M.; Wadsworth, M.H., 2nd; Treacy, D.; Trombetta, J.J.; Rotem, A.; Rodman, C.; Lian, C.; Murphy, G.; et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **2016**, *352*, 189–196. [[CrossRef](#)]
18. Gao, R.; Bai, S.; Henderson, Y.C.; Lin, Y.; Schalck, A.; Yan, Y.; Kumar, T.; Hu, M.; Sei, E.; Davis, A.; et al. Delineating copy number and clonal substructure in human tumors from single-cell transcriptomes. *Nat. Biotechnol.* **2021**, *39*, 599–608. [[CrossRef](#)]
19. Fan, J.; Lee, H.O.; Lee, S.; Ryu, D.E.; Lee, S.; Xue, C.; Kim, S.J.; Kim, K.; Barkas, N.; Park, P.J.; et al. Linking transcriptional and genetic tumor heterogeneity through allele analysis of single-cell RNA-seq data. *Genome Res.* **2018**, *28*, 1217–1227. [[CrossRef](#)]
20. Liu, X.; Griffiths, J.I.; Bishara, I.; Liu, J.; Bild, A.H.; Chang, J.T. Phylogenetic inference from single-cell RNA-seq data. *Sci. Rep.* **2023**, *13*, 12854. [[CrossRef](#)]
21. Kharchenko, P.V.; Silberstein, L.; Scadden, D.T. Bayesian approach to single-cell differential expression analysis. *Nat. Methods* **2014**, *11*, 740–742. [[CrossRef](#)]
22. Behm, M.; Ohman, M. RNA Editing: A Contributor to Neuronal Dynamics in the Mammalian Brain. *Trends Genet.* **2016**, *32*, 165–175. [[CrossRef](#)]
23. Gommans, W.M.; Mullen, S.P.; Maas, S. RNA editing: A driving force for adaptive evolution? *Bioessays* **2009**, *31*, 1137–1145. [[CrossRef](#)]
24. Ryu, D.; Kim, S.J.; Hong, Y.; Jo, A.; Kim, N.; Kim, H.J.; Lee, H.O.; Kim, K.; Park, W.Y. Alterations in the Transcriptional Programs of Myeloma Cells and the Microenvironment during Extramedullary Progression Affect Proliferation and Immune Evasion. *Clin. Cancer Res.* **2020**, *26*, 935–944. Erratum in *Clin. Cancer Res.* **2020**, *26*, 5049. [[CrossRef](#)] [[PubMed](#)]
25. Ji, A.L.; Rubin, A.J.; Thrane, K.; Jiang, S.; Reynolds, D.L.; Meyers, R.M.; Guo, M.G.; George, B.M.; Mollbrink, A.; Bergenstrahle, J.; et al. Multimodal Analysis of Composition and Spatial Architecture in Human Squamous Cell Carcinoma. *Cell* **2020**, *182*, 497–514.e22. [[CrossRef](#)] [[PubMed](#)]
26. Kiran, A.M.; O’Mahony, J.J.; Sanjeev, K.; Baranov, P.V. Darned in 2013: Inclusion of model organisms and linking with Wikipedia. *Nucleic Acids Res* **2013**, *41*, D258–D261. [[CrossRef](#)]
27. Gagnidze, K.; Rayon-Estrada, V.; Harroch, S.; Bulloch, K.; Papavasiliou, F.N. A New Chapter in Genetic Medicine: RNA Editing and its Role in Disease Pathogenesis. *Trends Mol. Med.* **2018**, *24*, 294–303. [[CrossRef](#)]
28. Patel, A.P.; Tirosh, I.; Trombetta, J.J.; Shalek, A.K.; Gillespie, S.M.; Wakimoto, H.; Cahill, D.P.; Nahed, B.V.; Curry, W.T.; Martuza, R.L.; et al. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* **2014**, *344*, 1396–1401. [[CrossRef](#)]
29. Ohishi, W.; Cologne, J.B.; Fujiwara, S.; Suzuki, G.; Hayashi, T.; Niwa, Y.; Akahoshi, M.; Ueda, K.; Tsuge, M.; Chayama, K. Serum interleukin-6 associated with hepatocellular carcinoma risk: A nested case-control study. *Int. J. Cancer* **2014**, *134*, 154–163. [[CrossRef](#)] [[PubMed](#)]
30. Jang, J.W.; Oh, B.S.; Kwon, J.H.; You, C.R.; Chung, K.W.; Kay, C.S.; Jung, H.S. Serum interleukin-6 and C-reactive protein as a prognostic indicator in hepatocellular carcinoma. *Cytokine* **2012**, *60*, 686–693. [[CrossRef](#)]
31. Yang, J.; Zeng, L.; Chen, R.; Zheng, S.; Zhou, Y.; Chen, R. Characterization of heterogeneous metabolism in hepatocellular carcinoma identifies new therapeutic target and treatment strategy. *Front. Immunol.* **2023**, *14*, 1076587. [[CrossRef](#)]
32. Jin, S.; Guerrero-Juarez, C.F.; Zhang, L.; Chang, I.; Ramos, R.; Kuan, C.H.; Myung, P.; Plikus, M.V.; Nie, Q. Inference and analysis of cell-cell communication using CellChat. *Nat. Commun.* **2021**, *12*, 1088. [[CrossRef](#)] [[PubMed](#)]
33. Cao, W.; Chen, Y.; Han, W.; Yuan, J.; Xie, W.; Liu, K.; Qiu, Y.; Wang, X.; Li, X. Potentiality of alpha-fetoprotein (AFP) and soluble intercellular adhesion molecule-1 (sICAM-1) in prognosis prediction and immunotherapy response for patients with hepatocellular carcinoma. *Bioengineered* **2021**, *12*, 9435–9451. [[CrossRef](#)]

34. Zhang, S.; Fang, W.; Zhou, S.; Zhu, D.; Chen, R.; Gao, X.; Li, Z.; Fu, Y.; Zhang, Y.; Yang, F.; et al. Single cell transcriptomic analyses implicate an immunosuppressive tumor microenvironment in pancreatic cancer liver metastasis. *Nat. Commun.* **2023**, *14*, 5123. [\[CrossRef\]](#)
35. Liu, F.; Huang, J.; He, F.; Ma, X.; Fan, F.; Meng, M.; Zhuo, Y.; Zhang, L. CD96, a new immune checkpoint, correlates with immune profile and clinical outcome of glioma. *Sci. Rep.* **2020**, *10*, 10768. [\[CrossRef\]](#)
36. Ye, W.; Luo, C.; Liu, F.; Liu, Z.; Chen, F. CD96 Correlates With Immune Infiltration and Impacts Patient Prognosis: A Pan-Cancer Analysis. *Front. Oncol.* **2021**, *11*, 634617. [\[CrossRef\]](#)
37. Denisenko, E.; de Kock, L.; Tan, A.; Beasley, A.B.; Beilin, M.; Jones, M.E.; Hou, R.; Muiri, D.O.; Bilic, S.; Mohan, G.; et al. Spatial transcriptomics reveals discrete tumour microenvironments and autocrine loops within ovarian cancer subclones. *Nat. Commun.* **2024**, *15*, 2860. [\[CrossRef\]](#) [\[PubMed\]](#)
38. Ross, E.M.; Markowetz, F. OncoNEM: Inferring tumor evolution from single-cell sequencing data. *Genome Biol.* **2016**, *17*, 69. [\[CrossRef\]](#) [\[PubMed\]](#)
39. Jahn, K.; Kuipers, J.; Beerenwinkel, N. Tree inference for single-cell data. *Genome Biol.* **2016**, *17*, 86. [\[CrossRef\]](#)
40. Zafar, H.; Tzen, A.; Navin, N.; Chen, K.; Nakhleh, L. SiFit: Inferring tumor trees from single-cell sequencing data under finite-sites models. *Genome Biol.* **2017**, *18*, 178. [\[CrossRef\]](#)
41. Zafar, H.; Navin, N.; Chen, K.; Nakhleh, L. SiCloneFit: Bayesian inference of population structure, genotype, and phylogeny of tumor clones from single-cell genome sequencing data. *Genome Res.* **2019**, *29*, 1847–1859. [\[CrossRef\]](#)
42. Vu, T.N.; Nguyen, H.N.; Calza, S.; Kalari, K.R.; Wang, L.; Pawitan, Y. Cell-level somatic mutation detection from single-cell RNA sequencing. *Bioinformatics* **2019**, *35*, 4679–4687. [\[CrossRef\]](#)
43. McCarthy, D.J.; Rostom, R.; Huang, Y.; Kunz, D.J.; Danecek, P.; Bonder, M.J.; Hagai, T.; Lyu, R.; HipSci, C.; Wang, W.; et al. Cardelino: Computational integration of somatic clonal substructure and single-cell transcriptomes. *Nat. Methods* **2020**, *17*, 414–421. [\[CrossRef\]](#) [\[PubMed\]](#)
44. Huang, A.Y.; Li, P.; Rodin, R.E.; Kim, S.N.; Dou, Y.; Kenny, C.J.; Akula, S.K.; Hodge, R.D.; Bakken, T.E.; Miller, J.A.; et al. Parallel RNA and DNA analysis after deep sequencing (PRDD-seq) reveals cell type-specific lineage patterns in human brain. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 13886–13895. [\[CrossRef\]](#)
45. Salehi, S.; Steif, A.; Roth, A.; Aparicio, S.; Bouchard-Cote, A.; Shah, S.P. ddClone: Joint statistical inference of clonal populations from single cell and bulk tumour sequencing data. *Genome Biol.* **2017**, *18*, 44. [\[CrossRef\]](#)
46. Jun, S.H.; Toosi, H.; Mold, J.; Engblom, C.; Chen, X.; O'Flanagan, C.; Hagemann-Jensen, M.; Sandberg, R.; Aparicio, S.; Hartman, J.; et al. Reconstructing clonal tree for phylo-phenotypic characterization of cancer using single-cell transcriptomics. *Nat. Commun.* **2023**, *14*, 982. [\[CrossRef\]](#)
47. Dou, J.; Tan, Y.; Kock, K.H.; Wang, J.; Cheng, X.; Tan, L.M.; Han, K.Y.; Hon, C.C.; Park, W.Y.; Shin, J.W.; et al. Single-nucleotide variant calling in single-cell sequencing data with Monopogen. *Nat. Biotechnol.* **2023**, *42*, 803–812. [\[CrossRef\]](#) [\[PubMed\]](#)
48. Muyas, F.; Sauer, C.M.; Valle-Inclan, J.E.; Li, R.; Rahbari, R.; Mitchell, T.J.; Hormoz, S.; Cortes-Ciriano, I. De novo detection of somatic mutations in high-throughput single-cell profiling data sets. *Nat. Biotechnol.* **2023**, *42*, 758–767. [\[CrossRef\]](#)
49. Picardi, E.; D'Erchia, A.M.; Lo Giudice, C.; Pesole, G. REDiportal: A comprehensive database of A-to-I RNA editing events in humans. *Nucleic Acids Res.* **2017**, *45*, D750–D757. [\[CrossRef\]](#) [\[PubMed\]](#)
50. Kravitz, S.N.; Ferris, E.; Love, M.I.; Thomas, A.; Quinlan, A.R.; Gregg, C. Random allelic expression in the adult human body. *Cell Rep.* **2023**, *42*, 111945. [\[CrossRef\]](#)
51. Trapnell, C.; Roberts, A.; Goff, L.; Pertea, G.; Kim, D.; Kelley, D.R.; Pimentel, H.; Salzberg, S.L.; Rinn, J.L.; Pachter, L. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **2012**, *7*, 562–578. Erratum in *Nat. Protoc.* **2014**, *9*, 2513. [\[CrossRef\]](#) [\[PubMed\]](#)
52. Satija, R.; Farrell, J.A.; Gennert, D.; Schier, A.F.; Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **2015**, *33*, 495–502. [\[CrossRef\]](#)
53. Ewing, A.D.; Houlahan, K.E.; Hu, Y.; Ellrott, K.; Caloian, C.; Yamaguchi, T.N.; Bare, J.C.; P'ng, C.; Waggott, D.; Sabelnykova, V.Y.; et al. Combining tumor genome simulation with crowdsourcing to benchmark somatic single-nucleotide-variant detection. *Nat. Methods* **2015**, *12*, 623–630. [\[CrossRef\]](#)
54. Qiu, X.; Hill, A.; Packer, J.; Lin, D.; Ma, Y.A.; Trapnell, C. Single-cell mRNA quantification and differential analysis with Census. *Nat. Methods* **2017**, *14*, 309–315. [\[CrossRef\]](#) [\[PubMed\]](#)
55. Tao, Z.; Wang, S.; Wu, C.; Wu, T.; Zhao, X.; Ning, W.; Wang, G.; Wang, J.; Chen, J.; Diao, K.; et al. The repertoire of copy number alteration signatures in human cancer. *Brief. Bioinform.* **2023**, *24*, bbad053. [\[CrossRef\]](#) [\[PubMed\]](#)
56. Kim, S.; Scheffler, K.; Halpern, A.L.; Bekritsky, M.A.; Noh, E.; Kallberg, M.; Chen, X.; Kim, Y.; Beyter, D.; Krusche, P.; et al. Strelka2: Fast and accurate calling of germline and somatic variants. *Nat. Methods* **2018**, *15*, 591–594. [\[CrossRef\]](#)

57. Wang, K.; Li, M.; Hakonarson, H. ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **2010**, *38*, e164. [[CrossRef](#)]
58. Lu, T.; Park, S.; Zhu, J.; Wang, Y.; Zhan, X.; Wang, X.; Wang, L.; Zhu, H.; Wang, T. Overcoming Expressional Drop-outs in Lineage Reconstruction from Single-Cell RNA-Sequencing Data. *Cell Rep.* **2021**, *34*, 108589. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.