Identification of Multi-landscape and Cell Interactions in the Tumor Microenvironment through High-Coverage Single-Cell Sequencing

Wenlong Zhong[1,3#], Ligang Wang[2#], Chengjunyu Zhang[1,3#], Tonglei Guo[2], Lihua Zhao[2], Daqin Wu[1,3], Fei Xie[2], Xiao Wang[2], Xiuxin Li[2], Fangxiao Wang[2], Minghui Li[2], Weiyue Gu[2]*, Tianxin Lin[1,3*], Xu Chen[1,3*]

1 Department of Urology, Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou, China.

2 Beijing Chigene Translational Medical Research Center Co. Ltd., Beijing, China.

3 Guangdong Provincial Key Laboratory of Malignant Tumour Epigenetics and Gene Regulation, Guangdong-Hong Kong Joint Laboratory for RNA Medicine, Sun Yat-Sen Memorial Hospital, Sun Yat-Sen University, Guangzhou, China.

To whom contributed this work equally

* To whom correspondence should be addressed. Weiyue Gu, E-mail: gwyue@163.com. Tianxin Lin, E-mail: lintx@mail.sysu.edu.cn. Xu Chen, E-mail: chenx457@mail.sysu.edu.cn.

Abstract

Single-cell RNA sequencing (scRNA-seq) is a widely used method for classifying cell types and states and revealing disease mechanisms. However, most contemporary scRNA-seq platforms fail to explore the multilandscape of RNA. Here, we designed a microfluidic chip that combines oligo-dT primers and Random Bridging Co-labeling (RBCL) RNA sequencing to develop an innovative Chigene scRNA-seq technology that can identify gene expression, mutations, and RNA splicing landscapes at the single-cell level. The Chigene scRNA-seq platform demonstrated exceptional performance, with minimal doublet rates of 0.94% (Chigene V1) and 1.93% (Chigene V2). Both versions exhibit high sensitivity, with Chigene V2 achieving nearly 100% RNA coverage and detecting over 1800 genes per cell on average. Targeted capture of single-cell gene mutations enhances mutation detection sensitivity. Moreover, this Chigene V2 platform has been validated in clinical samples for its ability to detect mutations, gene fusions and alternative splicing. The reliability of the platform was further corroborated via known functional gene mutation (CDKN1A) and fusion (FGFR3-TACC). To validate this method's potential for discovering novel gene mutations in clinical samples, our investigation revealed an intriguing cell subpopulation carrying an ARHGAP5 mutation in urothelial carcinoma. These cells exhibited high-frequency mRNA splicing and exhibited specific crosstalk with T cells, distinguishing them from the subpopulation with the ARHGAP5 wild-type phenotype. Overall, this method provides a robust scRNA-seq platform suitable for comprehensive analyses of clinical specimens at different genetic information levels, thereby offering significant potential in the discovery of novel genes and interactions at the single-cell level.

Keyword: Single-cell sequencing; High-Coverage; Urologic Tumors; Multi-landscape; Random Bridging Co-labeling

Introduction

Single-cell RNA sequencing (scRNA-seq) has revolutionized the field of tumor research by enabling the identification and characterization of cell types, states, lineages, and circuitry at the single-cell level[1-3]. These findings have greatly enhanced our understanding of tumor heterogeneity and the immune microenvironment. Genetic mutations[4], gene fusions, and alternative RNA splicing[5] are critical factors that contribute to tumorigenesis and tumor progression. These molecular alterations can significantly impact cellular processes, including cell proliferation, differentiation, metastasis and treatment response[6-9]. Integrating transcriptomic analysis with gene mutation data from whole exome sequencing (WES) and RNA alternative splicing data at the single-cell level holds immense potential for advancing the understanding of tumors. Although state-of-the-art scRNA-seq methods are highly sensitive and accurate for quantifying and characterizing cell states[10-13], most of these methods rely on barcoded oligo-dT primers that hybridize to polyadenylated transcripts for RNA capture and cDNA synthesis. This approach primarily detects short fragments located near the poly(A) tail or the 5' end of the transcript, leaving the remaining sequences of polyadenylated RNA molecules undetected. As a result, the accurate detection of single nucleotide variants (SNVs), gene fusions and alternative splicing events is limited[14-17]. Recently, several high-throughput scRNA-seq platforms for total-transcriptome full-length sequencing have emerged [18-20]. However, these methods pose limitations in terms of sequencing depth and mRNA detection sensitivity when total RNA is targeted within a fixed amount of data.

To address the challenges and fulfil the needs of scientific research and clinical practice, we developed a microfluidics-based high-sensitivity and full-length mRNA sequencing method termed Chigene-seq. For Chigene-seq, we employed oligo-dT primers for specific capture of mRNA during reverse transcription, followed by synthesis and specific labelling of the second strand via bridge oligo sequences on the first strand. Subsequently, amplification and sequencing were performed. Additionally, we designed targeted capture steps for specific mutations and gene fusions to further improve the detection efficiency in single cells. Furthermore, we developed a simplified and cost-effective 3' sequencing approach by using oligo-dT primers only. To validate the performance and clinical applicability of our single-cell sequencing platform, multiple cell

lines and human urological cancer samples were utilized. The results indicated the robustness of our platform and allowed for the exploration of gene mutations, gene fusions, and alternative splicing at the single-cell level. Moreover, the application value of this approach was demonstrated by the discovery of novel functional gene mutations in clinical samples. In summary, our novel Chigene scRNA-seq platform offers potential advantages for specific applications, such as mutation detection and alternative splicing analysis, providing expanded methods for discovering new functional genes in clinical samples.

Results

Overview of the Microfluidic Chip-based Random Bridging Co-labelling RNA Sequencing Method for Full-length Coverage

We developed a versatile workflow that enables the amplification and fragmentation of mRNA via microbeads carrying oligo-dT primers and random bridging primers on a microfluidic platform (Figure S1). This approach allows the detection of various natural barcodes, including somatic mutations, fusion genes, and RNA isoforms. Based on whether random bridging primers were utilized, the workflow was further divided into two versions: Chigene V1 (conventional 3' end sequencing) (Figure 1A-B) and V2 (mRNA full-length transcriptome sequencing) (Figure 1C). For Chigene V2, to further increase the detection sensitivity of gene mutation and fusion in single cells, we employed a targeted mutation library enrichment strategy with specific primers. In brief, the core process of this technology is illustrated in Figure 1D. First, the mRNA from single cells is captured on oligo microbeads that contain a 'cellular barcode sequence + polyT' structure, utilizing the polyA tail of the mRNA. Then the full-length first-strand cDNA was synthesized by reverse transcription. Subsequently, we added a fixed sequence to the end of cellular barcode oligos on the microbeads (as indicated by the three green stars in the legend of Figure 1D) which was unbound mRNA, allowing them to anneal with the bridge oligos that contained random sequences. These bridge oligos serve dual functions: the 5' end binds and ligates to the modified cellular barcode oligo through complementary pairing, while the 3' end anneals to the first-strand cDNA at multiple random regions via the random N sequence. Under the action of DNA polymerase, the synthesis of the second-strand DNA is anchored by the bridge oligos, ultimately leading to the systematic incorporation of identical cellular barcode sequences at multiple positions within the cDNA

99 molecules, thereby constructing a dual-strand cDNA library that features systematic single-cell
100 labeling characteristics. This random multipoint bridging strategy locks in cellular origin
101 information at the single-molecule level through the formation of covalent bonds. The bridged and
102 extended products were then amplified to generate sufficient cDNA for library construction and
103 subsequent next-generation sequencing (Figure 1C). To integrate the single-cell mRNA full-length
104 gene expression profiles with the genomic features of interest, we called the molecular features of
105 interest on the basis of genomic features detected by whole exome sequencing (WES) from the
106 same sample. Specifically, we enriched the library for the mutant and fusion genes of interest from
107 the transcriptome library via PCR with specific primers (Figure 1C). By extracting cell barcode
108 and transcript information from the sequencing data, we improved the detection rate of lowly
109 expressed gene mutations. This allowed us to explore the biological functions of cell
110 subpopulations with specific genomic features and their interactions with surrounding cells.
111
112 **Development and validation of the Chigene V1 platform on a microfluidic chip**
113 To evaluate the performance of the microfluidic chip for single-cell sequencing, the Chigene
114 V1 platform, which relies solely on chips and poly T primers, was first developed. To assess the
115 cell capture efficiency and doublet rate of the microfluidic chip, different quantities of HeLa cells
116 ranging from 5000 to 50000 cells were loaded. Microscopic counting revealed that the cell capture
117 efficiency was generally greater than 50%, and the doublet rate was mostly less than 5% (Figure
118 2A-B). Observations of the microbeads after loading revealed micropores with both single cells
119 and beads (Figure 2C). The time required for cross-contamination between micropores is
120 approximately 5 minutes, whereas cell lysis within the chip is achieved in approximately 12
121 seconds[21]. After the microbeads were loaded, cell lysis, RNA capture, and two quick washes were
122 performed within two minutes to minimize cross-contamination. The recovery of a large portion
123 of the microbeads (~70%) was achieved (Figure 2D). To evaluate the fidelity of the Chigene V1
124 platform, a standard mixed-species experiment was conducted using cultured human (A549) and
125 mouse (MC38) cell lines. A mixture of 2500 A549 cells and 2500 MC38 cells was profiled via
126 Chigene V1-seq. The analysis identified 2239 high-quality unique cell barcodes, indicating clear
127 separation of true cells from background noise (Figure 2E). The doublet rate was low at 0.94%
128 (Figure 2F), which is lower than that reported for other scRNA-seq methods (1.6--3.1%), such as

129 Seq-Well[22] and VASAdrop[19]. A more detailed comparison of doublet rates with other single-cell
130 platforms can be found in Supplementary Table 2. Additionally, high species specificity of UMI
131 (~97%) was observed, indicating that high-fidelity single-cell libraries were produced via the
132 Chigene V1 platform (Figure 2G). UMI and gene count distributions revealed that the Chigene V1
133 platform captured a median of 20,339 UMIs and 4672 genes in single A549 cells, with an average
134 of ~38k reads, and 17256 UMIs and 4805 genes in single MC38 cells, with an average of ~30k
135 reads (Figure 2H-I). These results demonstrated the high sensitivity of the Chigene V1 platform.
136 Next, we evaluated the ability of the Chigene V1 platform to distinguish cell populations in
137 complex primary samples. This was achieved by comparing its performance with that of 10x
138 Genomics, BD Rhapsody, and Singleron using human peripheral blood cells following erythrocyte
139 lysis. Compared with these mainstream scRNA platforms, Chigene V1-seq exhibited very high
140 cluster consistency (Figure 2J and Figure S2A-D) and similar cell proportions (Figure 2K-L and
141 Figure S2E). However, the performance of the Chigene V1 platform was more similar to that of
142 the BD Rhapsody and Singleron platforms because the same cell capture strategy (microfluidic
143 chip) was used. Furthermore, when a more complex urachal carcinoma sample was used for
144 performance comparison between Chigene V1 and 10x Genomics, similar results were observed
145 (Figure 2M-O, Figure S2F-H).
146
147 **Development of the Chigene V2 platform based on the Chigene V1 platform and Random**
148 **Bridging Co-labeling technique**
149 To enhance the capture of full-length mRNAs, we integrated the Random Bridging Co-
150 labelling (RBCL) technique into the Chigene V1 platform and developed the Chigene V2 platform.
151 The gene body coverage of the Chigene V2 platform was substantially greater than that of the
152 Chigene V1 and 10x platforms (Figure 3A). Moreover, species-mixing experiments were
153 conducted to analyse specificity on the Chigene V2 platform. The results revealed acceptable
154 doublet rates (1.93%) and specificity (~95%) in comparison to those of the Chigene V1 platform
155 (Figure 3B-C). Owing to its ability to capture full-length genes, compared with the Chigene V1
156 platform, the Chigene V2 platform resulted in a decrease in average UMIs and gene counts per
157 cell at comparable sequencing depths (~36k reads per human cell and ~26k reads per mouse cell)
158 (Figure 3D-E). Additionally, the Chigene V2 platform demonstrated more homogeneous coverage

159 of gene body regions than did other high-throughput full-length single-cell RNA sequencing
160 platforms, such as VASA-seq, snRandom, and bulk RNA-seq (Figure 3F). Compared with the
161 canonical full-length scRNA-seq assays (Smart-seq2), Chigene V2 exhibited similar gene body
162 coverage uniformity (Figure S3A). Gene body coverage analysis of the GAPDH and ACTB genes
163 revealed that unlike poly(A)-based 10x Genomics, which exhibited pronounced bias toward the 3'-
164 end, Chigene V2, VASA-seq, and snRandom did not display significant 3'- or 5'-end bias across
165 the gene body. As expected, Chigene V2 displayed greater coverage depth, which was attributable
166 to its emphasis on capturing mRNA within a comparable sequencing data volume (Figure 3G-H).
167 The impressive ability of the Chigene V2 platform to detect full-length mRNAs suggests that
168 it may be an ideal method for detecting mutations, gene fusions and alternative splicing at the
169 single-cell level. To improve the sensitivity of detecting mutations, particularly in low-abundance
170 transcripts and rare cells, a nested PCR method was employed for target enrichment of gene
171 mutations. To test this approach, a mixture of 293T, A549, and HeLa cell lines in equal
172 proportions was analysed via Chigene V2-seq to detect SNP sites in two housekeeping genes
173 (ACTB c.565G and GAPDH c.463A). The findings revealed the accurate identification of distinct
174 cell clusters representing the three cell lines, suggesting minimal cross-contamination associated
175 with the Chigene V2 platform (Figure 3I-J). Moreover, the detection of these two SNP sites in the
176 majority of cells underscores the strength of the Chigene V2 platform in conjunction with nested
177 PCR for mutation detection at the single-cell level (Figure 3K-L). Furthermore, by analysing
178 mixtures of non-small cell lung cancer cell lines containing specific mutations (EGFR_p. T790M
179 and KRAS_p. G12S) in varying proportions (Figure S3B-C and supplementary table 2), as well as
180 in a cohort of tumor samples (Figure S3D and supplementary table 3), we found that the detection
181 results closely aligned with the true distributions. Notably, in the Chigene V2 cell line mixing
182 experiments and the Chigene V2 detection of gene mutations in solid tumors, the results indicate
183 that the platform is capable of detecting rare cells, with a detection threshold of approximately 1%
184 or higher (Figure S3E-I). To evaluate whether the Chigene V2 platform could provide sufficient
185 information from clinical tissue samples similar to the Chigene V1 platform, we collected fresh
186 human urachal carcinoma samples and compared their RNA profiles via Chigene V1-seq and
187 Chigene V2-seq. The results revealed a strong correlation between the total RNA profiles obtained
188 via Chigene V1-seq and those obtained via Chigene V2-seq (Spearman R: 0.86, Figure 3M).

189 Furthermore, the majority of cell types were consistently identified by both methods in the same
190 sample (Figure 3N and Figure S3J-K). Unsupervised clustering analysis identified ten types of
191 cells in the merged sequencing data (Figure 3O and Figure S3L). No significant differences in the
192 proportions of cell types were detected between the V2-seq and V1-seq results (Figure 3P).
193
194 **Identification of multi-landscape by Chigene V2 platform in the clinical samples**
195 To assess the performance of the Chigene V2 platform in exploring alternative gene splicing,
196 mutation, and fusion, fresh human urothelial cancer samples were subjected to Chigene V2-seq,
197 whole-exome sequencing (WES) and bulk RNA-seq. A total of 4452 cells were identified from the
198 single-cell suspension, with an average of 1696 genes and 4665 UMIs detected per cell. These
199 cells were clustered into seven distinct cell types (Figure 4A and Figure S4A). By profiling full-
200 length transcripts via the Chigene V2-seq platform, alternative splicing (AS) patterns across
201 different cell types were identified through the quantification of inclusion rates of nonoverlapping
202 exonic regions. In this sample, seven different AS events (total: 3495 events) were detected
203 (Figure 4B), mainly in epithelial cells and T cells (Figure 4C). Notably, alternative first exon (AFE)
204 events were frequently detected (Figure 4B-C), further confirming the high coverage of mRNAs in
205 the Chigene V2 platform. Using the combined WES and RNA-seq data, ten genes with relatively
206 high mutation frequencies at both the DNA and RNA levels were selected for targeted enrichment
207 via the Chigene V2 platform (Figure S4C). UMAP visualization revealed a multitude of gene
208 mutations and comutations in epithelial cells (Figure 4D). To substantiate the reliability of the
209 Chigene V2 platform and investigate the functions of the identified mutated genes at the single-
210 cell level, particular attention was given to the CDKN1A p.F53Ffs*10 mutation (Figure 4E).
211 Compared with those in CDKN1A wild-type epithelial cells, gene ontology (GO) and KEGG
212 enrichment analyses revealed significant enrichment of upregulated genes related to cell cycle-
213 related biological processes (Figure 4F) and signalling pathways (Figure 4G) in CDKN1A-
214 mutated epithelial cells. Cell communication analysis revealed that, compared with wild-type cells,
215 CDKN1A-mutated epithelial cells had greater interactions with endothelial cells (Figure 4H).
216 Specifically, CDKN1A-mutated epithelial cells presented increased VEGF signalling to
217 endothelial cells (Figure 4I), suggesting a potential role in promoting angiogenesis. Similar

218 conclusions supporting the single-cell findings were obtained via TCGA data analysis (Figure
219 S4D-E).

220 Next, we explored the fusion gene in the Chigene V2 platform and identified the fusion gene
221 FGFR3-TACC in the single-cell dimension (Figure 4J and Figure S4B). GO enrichment analysis
222 revealed that downregulated genes in FGFR3-TACC fusion epithelial cells were significantly
223 enriched in antigen processing and presentation-related biological processes (Figure 4K),
224 potentially indicating a contribution to immune evasion. Cell communication analysis revealed
225 greater signal output from FGFR3-TACC fusion epithelial cells than from wild-type FGFR3
226 epithelial cells (Figure 4L). Ligand–receptor pair analysis revealed that FGFR3–TACC fusion
227 epithelial cells preferentially transmitted signals to both immune cells and nonimmune cells via
228 immunosuppressive ligand–receptor pairs and promoted angiogenesis and tumor proliferation
229 ligand–receptor pairs (Figure 4M–N), such as SEMA4D–PLXNB1/2[23-27],
230 COL6A1–SDC1–CD44[28], MPZL1–MPZL1[29-31], and WNT7A–FZD6/10+LRP5) [32]. These
231 findings suggest that the FGFR3-TACC fusion might hinder the efficacy of immunotherapy and
232 contribute to poor prognosis by affecting both immune cells and nonimmune cells. Similar
233 conclusions supporting the single-cell findings were obtained via TCGA data analysis (Figure
234 S4F-G). In summary, the Chigene V2 platform has the ability to explore alternative splicing,
235 mutations, and fusions at the single-cell level, and the results obtained are robust.

236

237 **Identification of novel mutated genes via the Chigene V2 platform**

238 To identify novel mutated genes, a fresh sample of urothelial carcinoma was analysed via
239 Chigene V2-seq. A total of 4491 cells with an average of 1821 genes and 4273 UMIs per cell were
240 identified. Unsupervised clustering revealed several distinct clusters representing different cell
241 types, including DCs, endothelial cells, epithelial cells, T cells, ILCs, and macrophages (Figure 5A
242 and Figure S5A). Analysis of the Chigene V2-seq data revealed multiple gene mutations in
243 epithelial cells (Figure 5B). Among these, we specifically focused on the ARHGAP5 gene, which
244 harbors the p.I307V mutation, owing to its high mutation frequency and the unclear biological
245 function and clinical significance associated with this alteration (Figure 5C). Sanger sequencing
246 validated the true existence of the mutation site in this gene (Figure S5B). GO enrichment and
247 alternative splicing analyses demonstrated a greater frequency of RNA splicing events in

248 ARHGAP5-mutated epithelial cells (Figure 5D-E). Cell communication analysis further revealed
249 that epithelial cells with mutated ARHGAP5 displayed a distinct number of interactions with other
250 clusters, particularly notable in their interaction with T cells (Figure 5F-G). Subsequent
251 examination of ligand−receptor pairs revealed notable differences between ARHGAP5-mutated
252 and wild-type tumor cells regarding the signals they transmitted to both immune and nonimmune
253 cells (Figure 5H). In contrast to ARHGAP5 wild-type epithelial cells, ARHGAP5-mutated
254 epithelial cells tend to transmit signals to dendritic cells (DCs) via the F11R–F11R interaction,
255 which can impede DC migration to lymph nodes and the initiation of tumor-specific immune
256 responses [33]. Additionally, in ARHGAP5-mutated epithelial cells, signals are specifically
257 transmitted to T cells through APP-CD74, which is known to exert an inhibitory effect on the
258 activation of CD4+ T cells (Figure 5I). Moreover, ARHGAP5-mutated epithelial cells also
259 transmit signals to nonimmune cells involved in tumorigenesis, metastasis, and angiogenesis
260 (Figure 5J) via MPZL1-MPZL1[29-31], F11R-F11R[34], and COL6A1-SDC1[28]. Consequently, it is
261 speculated that urothelial carcinoma patients with ARHGAP5 mutations may exhibit a poorer
262 response to immunotherapy. Additionally, TCGA data analysis (Figure S5C) and validation
263 through cell lines (Figure S5D) supported the findings related to the ARHGAP5 mutation. Overall,
264 our experimental results corroborate the findings from the TCGA analysis and provide further
265 validation of the role of the ARHGAP5 p.I307V mutation in promoting pathways that enhance
266 tumor malignancy, metastatic potential and immune evasion. Furthermore, these findings validate
267 the functionality of our Chigene V2 platform, demonstrating its ability to provide insights into the
268 functional implications of this mutation at the single-cell level, underscoring its importance in
269 tumor biology.

270 To gain a more comprehensive understanding of the prevalence of ARHGAP5 mutations
271 across various cancers, we searched the TCGA database (Figure S5E). Intriguingly, our analysis
272 revealed a mutation frequency of approximately 2.6% for ARHGAP5 in urothelial carcinoma
273 (Figure S5F). To assess the clinical implications of these mutations, propensity score matching
274 (1:1) was conducted, matching patients with ARHGAP5 mutations to their wild-type counterparts
275 based on variables such as cancer type, age, sex, race, and pathological stage (Figure S5G).
276 Despite not reaching statistical significance, an intriguing pattern emerged: progression-free
277 survival (PFS) tended to be shorter in the mutant patients (Figure S5H). The associations between

ARHGAP5 mutations and immunotherapy response in cohorts of patients who underwent immunotherapy for urothelial carcinoma[35] or advanced clear cell renal cell carcinoma[36] were further investigated. Interestingly, no ARHGAP5 mutations were detected among the 9 responders in the urothelial cancer cohort, whereas one such mutation was present among the 16 nonresponders. In the advanced clear cell renal cell carcinoma cohort, ARHGAP5 mutations were identified in 5 out of 261 patients. Among these patients, 2 had stable disease (SD), and 3 experienced progressive disease (PD). The distribution of these clinical responses indicates a possible association between ARHGAP5 mutations and unfavourable treatment outcomes. By integrating data from the TCGA database, immunotherapy cohorts, and WES analysis, we propose that ARHGAP5 gene mutations might be linked to decreased progression-free survival (PFS) and poor response to immunotherapy.

Discussion

In this study, we introduced a microfluidic chip-based high-throughput single-cell RNA sequencing (scRNA-seq) system designed for comprehensive analysis of mRNA expression, mutations, alternative splicing, and even gene fusions at the single-cell level. Our scRNA-seq platform consists of two versions. Compared with mainstream platforms such as 10X Genomics, BD Rhapsody, and Singleron, Chigene V1, a conventional single-cell sequencing platform utilizing 3' end sequencing, has demonstrated equivalent performance and scalability. Chigene V2 specifically captures mRNAs with more uniform coverage across the gene body and achieves greater sequencing depth for coding sequences. Notably, our Chigene scRNA-seq platform employs standard molecular biology procedures and a well-established microwell barcoding platform. As a result, the Chigene scRNA-seq platform is user friendly and can be readily used for large-scale applications.

Most contemporary single-cell RNA sequencing (scRNA-seq) platforms depend primarily on the hybridization of barcoded oligo-dT primers. However, this approach limits the ability of these methods to detect short fragments (~400--600 base pairs) near the poly(A) tail, thereby hampering the thorough investigation of gene mutations, gene fusion, and alternative splicing at the single-cell level. Recent advancements in total RNA full-length scRNA-seq methods have addressed these challenges [18, 20, 37]. However, the use of total RNA sequencing techniques presents inherent

limitations in sequencing depth dedicated to mRNAs owing to the presence of noncoding RNAs and alternative splicing variants in the total RNA pool. Consequently, these molecules can compete for sequencing data, resulting in a reduced sequencing depth for mRNA detection. The Chigene V2 platform introduces a novel random bridging colabelling technique, facilitating the specific capture of full-length mRNAs. This method demonstrated improved alignment with conventional 3' single-cell sequencing, effectively minimizing interference from rRNA and other transcripts. This innovation eliminates the need for additional bioinformatics or enrichment methods to remove rRNA, thereby saving time and experimental costs. Furthermore, data utilization can be enhanced by increasing the sequencing depth for mRNAs within the same data volume. In addition, by incorporating targeted capture techniques, our approach facilitates the detection of low-abundance mutations and fusions, which substantially enhances sensitivity (Figure S6A-B). This combination of specific mRNA capture and targeted mutation detection results in notable improvements in both the breadth and sensitivity of our platform, allowing for a more comprehensive analysis of mutations and fusions within the transcriptome. These innovations underscore the novelty and effectiveness of our methods in advancing transcriptome-level research.

The theoretical throughput of Chigene V2 is equivalent to that of conventional 3' single-cell sequencing. Among the existing single-cell sequencing platforms, the Chigene V2 platform is superior for mRNA analysis because of its more even gene body coverage and sequencing depth. These advantages enhance the analysis of alternative splicing in single-cell polyadenylated RNA species and facilitate the identification of rare subpopulations for precision diagnosis and treatment of human diseases. Moreover, when combined with targeted capture, the Chigene V2 platform achieves increased sensitivity for detecting gene mutations and gene fusions. These capabilities enable the simultaneous exploration of gene expression, gene mutations, alternative splicing, and gene fusion at the single-cell level.

To validate the potential of our platform for multidimensional clinical sample testing, we utilized our Chigene V2 platform to concurrently detect mutations, alternative splicing, and gene fusions in several human cancer samples. When we investigated known functional gene mutations, such as CDKN1A, and gene fusions, such as FGFR3-TACC, via the Chigene V2 platform, the results implied that enrichment analyses of differential gene expression between wild-type and

338 mutant cells were consistent with previously reported functional changes in tumor cells.
339 Additionally, our analysis revealed that intercellular signalling can affect the TME, which
340 deepened our understanding of genes with known functions. Notably, in this study, we identified a
341 novel gene mutation site in ARHGAP5, which has not been previously reported in the context of
342 tumors. We utilized this example to illustrate the finer resolution of our Chigene V2 platform in
343 bladder cancer research. Importantly, through TCGA database analyses and experimental
344 validation via cell line studies, we further demonstrated that ARHGAP5 functions as a tumor
345 suppressor gene, potentially regulating tumor cell malignancy and immune evasion. Additionally,
346 analyses from the TCGA database indicate that this mutation has potential as a prognostic
347 biomarker across various cancers (Figure S5). Interestingly, more recently, ARHGAP5 has also
348 been implicated as a tumor suppressor in colorectal cancer, where its mutation is associated with
349 shortened progression-free survival (PFS)[38]. These findings reinforce the potential of our single-
350 cell findings to reveal novel gene mutations and their roles in cancer. We plan to conduct more
351 detailed mechanistic studies and animal model validations to assess the impact of the ARHGAP5
352 mutation on bladder cancer comprehensively.

353 Nevertheless, it is important to acknowledge certain limitations of our technique. First, like
354 other full-length transcriptome sequencing technologies, Chigene V2 currently supports the
355 detection of mutations only in expressed genes. Mutations present in non-expressed genes
356 currently remain undetectable. However, it is worth reconsidering whether these nonexpressed
357 mutations may nonetheless influence tumor development and progression, as this limitation may
358 not have a profound impact on the overall applicability of our platform. Second, the single-cell
359 cDNA amplification method relies on poly(T) primers, which restrict the detection of mRNA
360 molecules lacking a poly(A) tail, such as histone mRNAs[39]. Consequently, these specific
361 transcripts were not captured or detected by our mRNA-seq assay. Moreover, given that our
362 technology captures a broader range of mRNA regions, the data volume required for analysis is
363 relatively greater than that of some existing platforms. This aspect should be considered when
364 evaluating the practicality of implementing Chigene V2 in various research settings. Finally, as
365 our assay utilizes mature mRNAs as templates, mutations present within intronic and intergenic
366 regions are not detected. These limitations could be addressed through future technical
367 refinements.

368 In conclusion, the straightforward experimental protocols and comprehensive transcriptomic
369 information offered by our platform show promising potential for the significant contributions of
370 Chigene scRNA-seq to large-scale applications in both basic and clinical research in the future.
371 The promotion and application of our platform are expected to advance the precision of
372 personalized medicine and therapeutic target discovery by providing deeper insights into the
373 genetic landscape of tumors.
374
375 **Method**
376 **Microfluidic chip device fabrication and microbeads**
377 The chip comprises a support layer, a chip microstructure layer, and a chip top cover layer. The
378 support layer was composed of rigid glass, whereas all other layers were constructed using
379 polydimethylsiloxane (PDMS) material with mold casting. For the chip microstructure layer, a
380 photomask was created based on the chip design, and a mold was fabricated via traditional
381 photolithography methods (with an SU-8 negative photoresist). The resulting mold consisted of an
382 array of cylindrical microstructures with dimensions of 52 μm (height) and 38 μm (diameter) and
383 a center-to-center distance of 48 μm between cylinders. The PDMS mixture, consisting of a
384 PDMS base and curing agent at a ratio of 10:1, was adequately mixed and poured onto the chip
385 microstructure layer mold. After baking at 80°C for 30 min, the cured PDMS chip microstructure
386 layer was gently removed from the mold. Moving to the chip top cover layer, a mold was prepared
387 on the basis of the chip design, including a chip chamber framework section and a top cover
388 section. Similarly, PDMS was cast onto the mold to form the chip top cover layer. To complete the
389 chip assembly, 1 mm diameter holes were punched in the top cover section, which served as the
390 inlet and outlet ports. The glass layer, PDMS chip microstructure layer, and PDMS chip top cover
391 layer were subsequently treated with plasma and bonded together to form the final microwell. The
392 microbeads were purchased from Chemgenes Corporation (MACOSKO-2011-10 B) and were
393 filtered through a 40 μm strainer. A detailed protocol for the Chigene V1 assay, including the oligo
394 list, is provided in the Supplementary Materials (Supplementary Methods and Supplementary
395 Table 1).
396 **Species mixture experiment and human samples**
397 HeLa cells, A549 cells, 293T cells, and MC38 cells were obtained from the CAMS Cell Culture

398 Center and cultured in Dulbecco's modified Eagle's medium (DMEM, Gibco, Cat #11965092)
399 supplemented with 10% heat-inactivated fetal bovine serum (FBS, Gibco, Cat # 26010074). The
400 cells were maintained at 37 °C in a 5% CO_2 incubator (Thermo Heracell 240i) and passaged every
401 2 days. For the species mixture experiment, A549 and MC38 cells were harvested and washed
402 three times in PBS by centrifugation at 4 °C and 600 × g for 3 min. Subsequently, the cells were
403 individually counted, and equal numbers of cells from each cell line were mixed. The resulting
404 cell mixture was subjected to single-cell RNA sequencing (scRNA-seq) following the Chigene
405 scRNA-seq protocol. The collection of human peripheral blood and urological cancer samples, as
406 well as the research conducted in this study, was approved by the Research Ethics Committee of
407 Sun Yat-sen Memorial Hospital, Sun Yat-sen University (approval numbers: SYSKY-2023-396-
408 01). Clinical information was collected after written informed consent was obtained.

409 **Library preparation**

410 The chip was first preprocessed using PBS I buffer (PBS + RNase inhibitor), and then
411 approximately 15,000 cells and capture beads were loaded onto the chip in turn. After the
412 redundant cells and beads were washed out, cell lysis occurred in the microwells of the chip. The
413 RNA molecules released from the lysed cells were captured by beads coated with oligo sequences
414 in the same microwell. Then, the lysis buffer in the chip was rapidly exchanged with air and high-
415 density wash buffer, and all the beads, including the mRNA-captured beads, were recovered by
416 washing. Approximately 80% of the beads were collected by two washes, and additional washing
417 steps helped to obtain more beads. Next, cDNA molecules with cell barcodes and UMI sequences
418 were obtained through reverse transcription. After ExoI digestion and amplification, the full-
419 length cDNA was enriched, and a single-cell RNA sequencing (scRNA-seq) library was
420 constructed via modified Tn5 transposase-based methods. Briefly, 5–10 ng of cDNA was
421 fragmented and amplified to construct the sequencing library. This approach will help to obtain
422 more information on the gene body for mRNA isoform analysis. Sequencing was performed via
423 PE150 bp on the MGI-T7 platform (MGI, China).

424 **Random bridging colabelling for Chigene V2-seq**

425 To obtain full-length mRNA, after reverse transcription and template switching, the mRNA-
426 free barcode oligo was annealed with the tag primer via the following reaction mixture: 100,000
427 microbeads in 35 μl of NF-water, 5 μl of 10 μM tag primer, and 10 μl of 20X SSC Buffer (NEB).

428 The reaction mixture was incubated at room temperature for 15 minutes. After annealing, the
429 microbeads were washed with PBS. To modify the 3' end of the barcode oligo, an extension
430 mixture consisting of 2 μl of NEB Buffer 10X, 5 μl of dNTPs, 2 μl of Klenow (exo-), and 11 μl of
431 NF-water was added. The mixture was then incubated at 37°C for 30 minutes. Denaturing buffer
432 (Illumina) was added to remove the mRNA and tag primer from the microbeads. The microbeads
433 were then washed twice with PBS. Bridge oligo1 and bridge oligo2 were annealed to form a
434 bridge dimer. One end of the bridge dimer was reverse complemented to the modified barcode
435 oligo, whereas the other end contained tens of random nucleotides "N" at the 3' tail. When the
436 modified barcode oligo annealed with the bridge dimer, the 3' end of the barcode oligo joins the
437 neighboring 5' end of the N-tail hand. The bridge capture mixture, which consisted of 10 μl of 20X
438 SSC buffer, 10 μl of 10 μM bridge dimer, and 100,000 microbeads in 30 μl of NF-water, was
439 incubated at room temperature for 15 minutes. The microbeads were then washed with PBS and
440 suspended in the ligation reaction mixture, which included 2 μl of 10X T4 DNA ligase buffer, 5 μl
441 of ATP, 2 μl of T4 DNA ligase, and 11 μl of NF-water. This mixture was incubated at room
442 temperature for 30 minutes. After washing with PBS, extension was performed via the following
443 reaction mixture: 2 μl of NEB Buffer 10X, 5 μl of dNTPs, 2 μl of Klenow (exo-), and 11 μl of NF-
444 water. To improve efficiency, repeating the steps from bridge capture to extension 2--3 times is
445 recommended. To block the free barcode oligo, ddNTPs were used, and then the cDNA was
446 amplified and purified.

447 **Targeted capture of specific regions of genes**

448 For the confirmation of mutation sites or gene fusions by WES, nested PCR primers were
449 designed upstream and downstream of each target region. The upstream-out primers, downstream-
450 out primers, upstream-in primers, and downstream-in primers were mixed separately. A total of
451 2.5 ng of cDNA was used as a template for amplification with the upstream and downstream
452 primers. Subsequently, secondary PCRs were carried out using the upstream and downstream
453 primers. After each round of amplification, the reaction products were individually purified via
454 magnetic beads. The purified products from each reaction were then combined. Index PCR was
455 performed on the combined products to generate a sequencing library. All the libraries were
456 subjected to quality control analysis via an Equalbit 1× dsDNA HS Assay Kit (Vazyme, China)
457 and a Standard (S2) Cartridge Kit (Bioptic, Taiwan, China). This quality control ensures that the

libraries meet the desired standards for further downstream analysis.

Data analysis

Preprocessing of the Chigene V1/V2-seq data. First, the raw read data were preprocessed via a Python script. For each Read1, the cell-specific barcode (16 nucleotides) located at the beginning of the read was identified. Hamming distance correction was applied by comparison to a whitelist of 100,000 barcodes provided by the Chigene platform. Barcodes within the whitelist or with a Hamming distance of 1 or less were considered valid cell barcodes and retained for downstream analysis. Barcodes not present in the whitelist or with a Hamming distance exceeding 1 were discarded to ensure accurate cell identification. The corrected barcode from Read1, along with the UMI (8 nucleotides), was appended to the read ID of Read2. The resulting reads were aligned to the reference genome (Chigene V1: GRCh38, Chigene V2: GRCh37, mixed human–mouse genome: GRCh38 and the GRCm39 hybrid genome) via the STAR aligner (version 2.7.1a) with default parameters. Reads uniquely mapped to the reference genome were assigned to genes via featureCount (version 2.0.1). UMItools (version 1.0.1) was used to correct UMI errors and perform UMI quantification within individual cells, resulting in the generation of the raw matrix. The EmptyDrops cell calling algorithm was utilized to identify true cells from the background and generate a gene expression matrix for subsequent analysis.

Clustering and downstream analysis. The single-cell expression matrix was imported into Seurat (version 3.0.2). Cells with fewer than 500 detected genes or more than 20% mitochondrial UMI counts were filtered out. Genes detected in more than 3 cells were retained. The genes were then ranked in descending order based on residual variance estimated via the "vst" method implemented in the FindVariableFeatures function of Seurat. The top 2000 genes, identified as highly variable genes (HVGs), were selected for further analysis. Principal component analysis (PCA) was performed on these highly variable genes. A shared nearest neighbor (SNN) graph was constructed using the top 15 principal components, and the cells were clustered via the Leiden algorithm. The resolution parameter for clustering was set to 0.6, while all other parameters were kept at the default values. To visualize the clusters, uniform manifold approximation and projection (UMAP) was applied in Seurat. After completing the clustering analysis, the cell types were annotated via a cell type (version 0.1.9).

Expression correlation analysis for the Chigene V1/V2 platform. To compare the gene

488 expression correlation between the Chigene V1 and Chigene V2 platforms, we imported the
489 expression matrices of both platforms into Seurat (version 3.0.2) and performed data
490 normalization and standardization using the same default parameters. We then utilized the Seurat
491 function “AverageExpression” to calculate the average expression levels of genes in each platform.
492 Finally, we used the “cor” function in R (version 4.0.2) to calculate the Spearman correlation
493 coefficient based on the natural logarithm of the average expression levels.

494 **Read coverage analysis of the gene body.** The public VASA-seq data can be accessed at the
495 Gene Expression Omnibus (GEO) under the accession code GSE176588. The public snRandom-
496 seq data can be obtained from the Genome Sequence Archive (GSA) with the accession code
497 "HRA003712". The 10x data were obtained from the PBMC data sequenced by Chigene via the
498 10x platform. The bulk RNA-seq data were obtained from transcriptome sequencing data. All the
499 data sets were aligned to the GRCh38 reference genome via the STAR Aligner (version 2.7.1a)
500 with the same parameters, resulting in BAM files. For all single cells, the BAM files were used as
501 bulk data. RSeQC (version 4.0.1) was used to calculate the read coverage over the gene body.

502 **Variant calling.** The barcode corrected from Read1, along with the UMI (8 nucleotides), was
503 added to the read ID of Read2. These reads were aligned to the GRCh37 reference genome via
504 STAR (version 2.7.1a) with default parameters. To remove duplicates, the umi_tools' group and
505 dedup functionalities were applied based on genomic coordinates and UMIs. This process helps
506 eliminate PCR amplification biases and improves the accuracy of variant detection. For variant
507 calling, BCFtools (version 1.9) was utilized to call variants at the single-cell level. The identified
508 mutations were subsequently annotated via ANNOVAR to obtain additional information regarding
509 mutation characteristics and potential functional impacts. Cells that harboured detected mutations
510 were classified as mutant cells, whereas cells without detected mutations were defined as wild-
511 type cells.

512 **Cell interaction analysis.** Cell communication analysis was performed via the R package
513 CellChat (version 1.5.0). After preprocessing the single-cell RNA sequencing data by applying
514 quality control measures and normalizing expression values, we utilized CellChat to identify
515 ligand–receptor pairs between different cell populations within the TME. The default parameters
516 provided by the CellChat package were employed for comprehensive coverage of potential cell
517 communication events. The probabilities of communication between cell types were estimated

518 based on the identified ligand–receptor pairs, which represent the likelihood of effective
519 information exchange. Signalling pathways involved in intercellular communication were inferred
520 via the built-in algorithms provided by CellChat. This comprehensive analysis aimed to elucidate
521 the intricate landscape of cell–cell interactions within the tumor microenvironment and shed light
522 on potential communication between different cell populations.

523 **GO and KEGG analyses.** For the enrichment analysis of DEGs between comparison groups, the
524 FindMarkers function in Seurat was used. This function allows the identification of differential
525 gene sets based on specified criteria such as the percentage of cells expressing the gene (pct) and
526 the p value (p_val). For example, genes with pct > 0.1 and p_val < 0.05 were selected. To perform
527 Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment
528 analyses of the DEGs, the R package clusterProfiler was used. This package provides functions for
529 functional enrichment analysis and visualization. By default, the package uses hypergeometric
530 tests for enrichment analysis and includes a set of predefined gene sets from databases such as GO
531 and KEGG.

532 **Statistical Analysis.** Statistical details for each experiment are provided in the figure legends. Pre-
533 processing of the Chigene V1/V2-seq data was described in the Data Analysis section. The single-
534 cell isolation experiment and doublet rate experiment were repeated more than three times
535 independently with similar results. The p value (p) for the Sperman's correlation coefficient (R)
536 was computed from a two-sided permutation test. The ggpubr package (0.4.0) in R was used for
537 statistical analysis.

538

539 **Supporting Information**

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

541

542 **Acknowledgements**

543 This study was supported by the National Key Research and Development Program of China
544 (Grant No. 2022YFC2408300), the National Natural Science Foundation of China (Grant No.
545 82322056, 82341018, 82072827, 81825016, and 82273421), the Guangdong Basic and Applied
546 Basic Research Foundation (Grant No. 2021B1515020009), the Science and Technology Program
547 of Guangzhou (Grant No. 2023A03J0718, 2024B03J1234, and 2024A04J6558), Fundamental

Research Funds for the Central Universities, Sun Yat-sen University (23ykbj002 for Xu Chen), the Guangdong Provincial Clinical Research Centre for Urological Diseases (2020B1111170006), and the Guangdong Science and Technology Department (2020B12060018, 2018B030317001, and 2017B030314026).

Conflicts of interest

Ligang Wang and Weiyue Gu have submitted a patent application to the Chinese patent office regarding the methodology described in this work (application number: CN 202311585591.9). Several authors are actively involved in the commercialization of the technique and are affiliated with Chigene, Inc. (Weiyue Gu is the founder; Ligang Wang, Tonglei Guo, Lihua Zhao, Minghui Li, Fei Xie, Xiao Wang, Xiuxin Li and Fangxiao Wang are employees). The remaining authors declare that they have no competing interests.

Authors' contributions

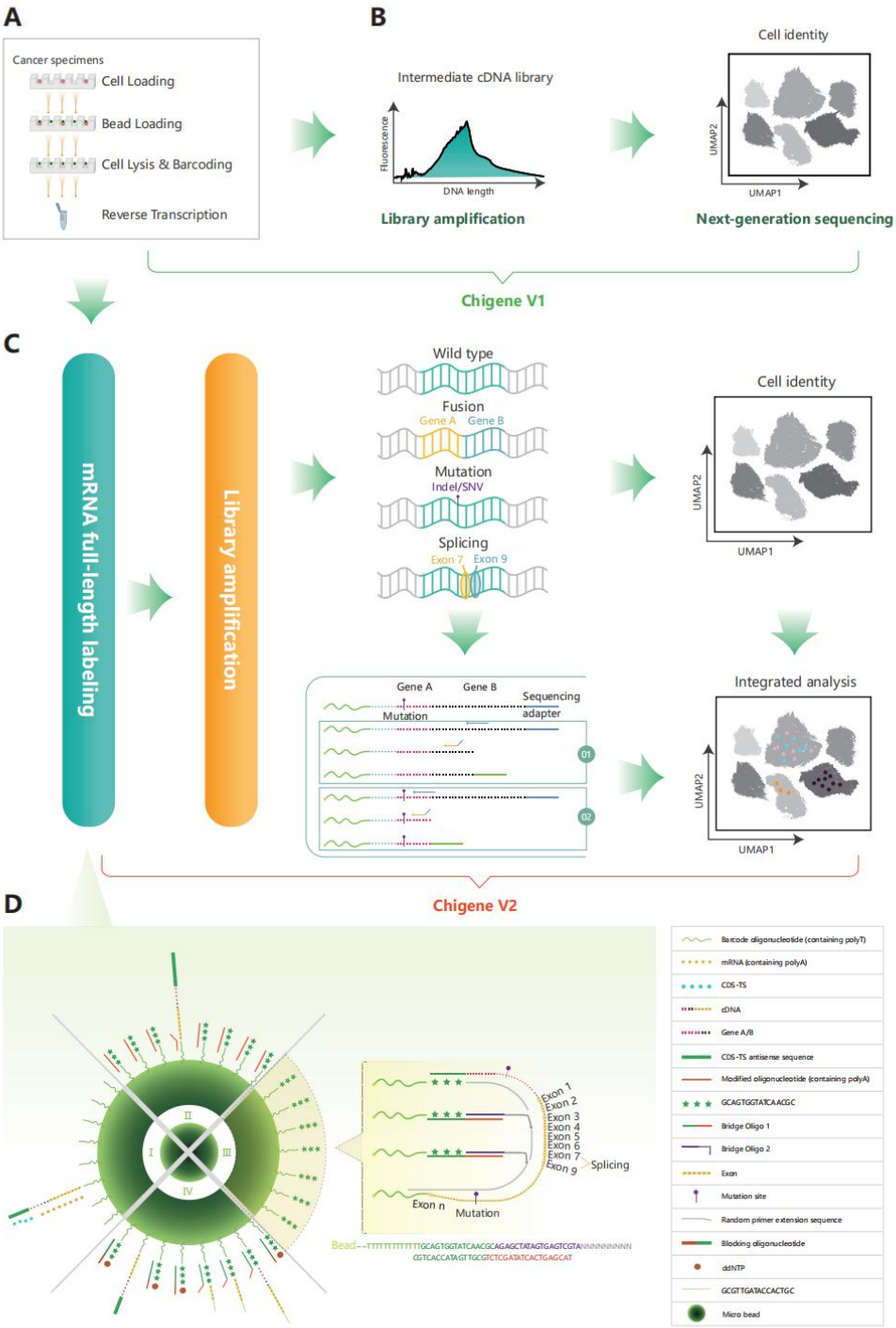
Weiyue Gu, Tianxin Lin, Xu Chen and Ligang Wang conceptualized the study. Ligang Wang, Xiao Wang, and Fangxiao Wang were responsible for the experimental design and execution. Tonglei Guo, Fei Xie, and Xiuxin Li conducted the bioinformatics analysis. Wenlong Zhong, Daqin Wu, and Xu Chen contributed to the clinical tissue selection and PBMC acquisition. Chengjunyu Zhang performed various cellular experiments, including ARHGAP5 gene knockdown and overexpression studies, to further investigate the functional role of ARHGAP5 in cancer biology. Lihua Zhao and Minghui Li performed the data analysis. Lihua Zhao contributed to the overall conceptualization and design of the manuscript and drafted the initial version. Lihua Zhao, Wenlong Zhong and Xu Chen participated in manuscript revisions and discussions related to clinical aspects. Weiyue Gu, Xu Chen and Tianxin Lin provided supervision throughout the project. All the authors have read and approved the final version of the manuscript.

Data availability

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

578 **Figures and Figure legends**

579



580

581 **Figure 1. Schematic workflow of Chigene-seq based on the microfluidic chip platform. (A)**
582 Unique mRNA capture was performed via oligo-dT on a microfluidic chip, and the recovered
583 microbeads were subjected to reverse transcription. The term “cancer specimens” refers to tumor

584 tissue samples obtained from patients diagnosed with urological cancer and collected through
585 biopsy procedures. **(B)** An intermediate cDNA library was amplified and subjected to next-
586 generation sequencing for conventional 3' scRNA-seq (Chigene V1). **(C)** Overview of the Chigene
587 V2 Library construction and analysis. The process begins with full-length mRNA labelling,
588 followed by library amplification. Different forms of genetic variation, including wild type, fusion
589 (gene A and gene B), mutation (indels/SNVs), and splicing (exon 7 and exon 9), are depicted to
590 highlight the library's ability to detect these events. The output is used for cell identity
591 determination and integrated analysis. The panel on the middle of Figure 1C provides a detailed
592 view of the sequencing output, indicating the relationships between genes, mutations, and the
593 corresponding sequencing adapters. Panel 01 captures fusion events, whereas Panel 02 captures
594 mutations. **(D)** Detailed schematic illustrating the rationale behind Chigene V2. The circular
595 layout presents the full-length mRNA labeling framework. Notably, the three green stars represent
596 the fixed sequence at the end of the cellular barcode oligos on the microbeads. The legend on the
597 right defines the various components of the schematic.

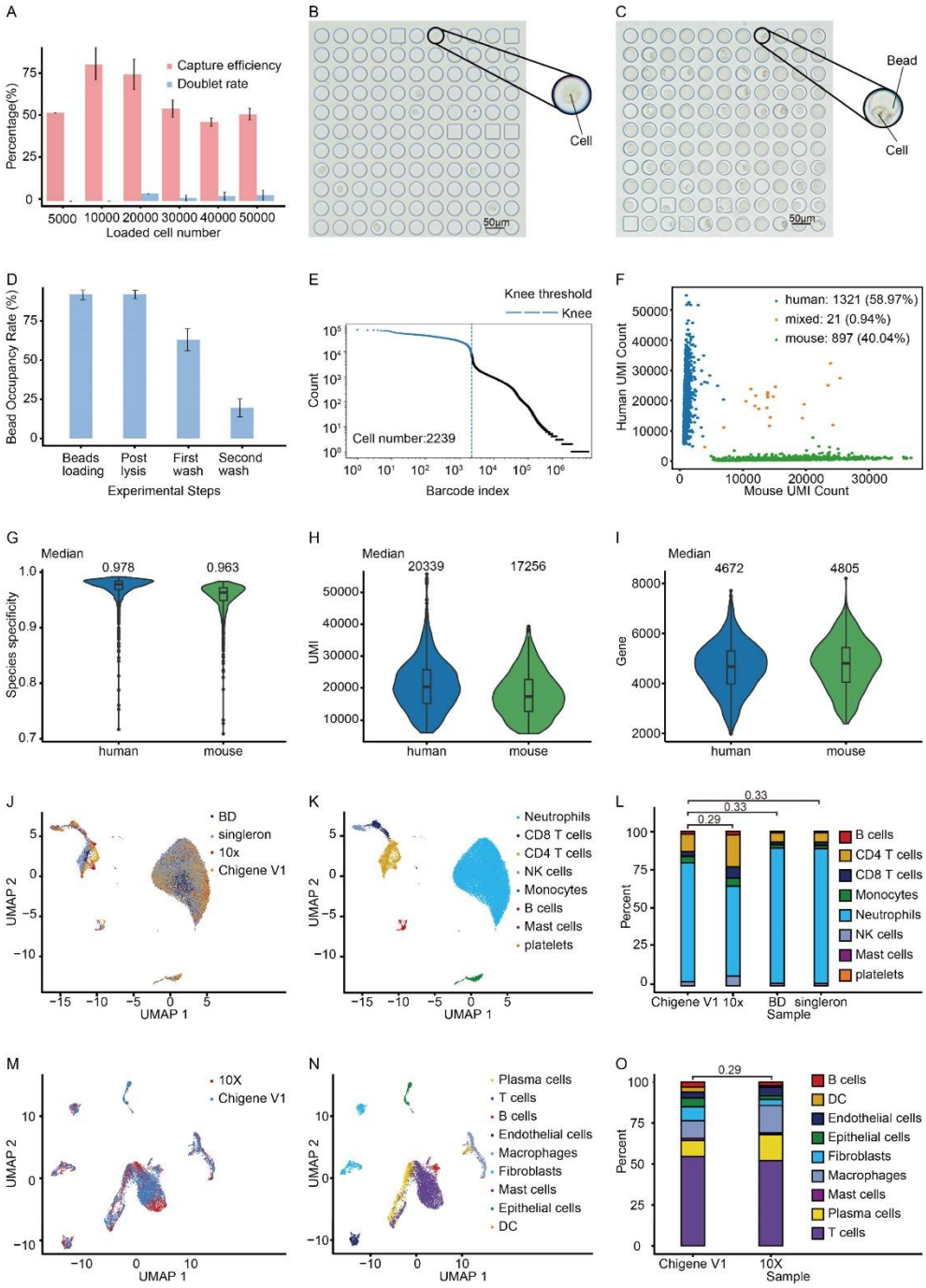


Figure 2. Development and validation of the Chigene V1 platform based on the microfluidic chip. (A) Capture rates and doublet rates for various numbers of loaded cells. **(B)** Bright-field images depicting a region of the device after cell loading. **(C)** Bright-field images showing a cell/bead-loaded device. **(D)** Bead occupancy rate following bead loading, cell lysis, and two rounds of washing. **(E)** Barcode plot used for identifying barcodes representing true cells (blue line). The barcodes of the A549-MC38 mixed cells were arranged in descending order on the basis

605 of gene count. **(F)** Species-mixing scatter plot showing the single-cell capture efficiency and
606 doublet rate of Chigene V1-seq. **(G)** Species specificity of UMIs in a human–mouse cell mixture.
607 The median species specificity of UMIs in human cells (A549) was 0.978. The median species
608 specificity of UMIs in mouse cells (MC38) was 0.963. **(H-I)** Violin plots and box plots showing
609 the number of UMIs and genes detected in individual A549 and MC38 cells. A549 cells: n = 1321;
610 MC38 cells: n = 897. The data depicted in the box plots correspond to the first and third quartiles
611 (lower and upper hinges) and the median (centre). **(J and K)** UMAPs displaying single-cell
612 profiles (dots) from PBMCs, with colors indicating the different platforms used (J) and each cell
613 type (K). **(L)** Proportion of cells of each cell type (y-axis) detected via different platforms (x-axis)
614 in the PBMC libraries. **(M and N)** UMAPs displaying single-cell profiles (dots) of human urachal
615 carcinoma cells, with colors indicating the different platforms used (M) and each cell type (N). **(O)**
616 Proportion of cells of each cell type (y-axis) detected via Chigene V1 and 10x Genomics (x-axis)
617 in human urachal carcinoma libraries.

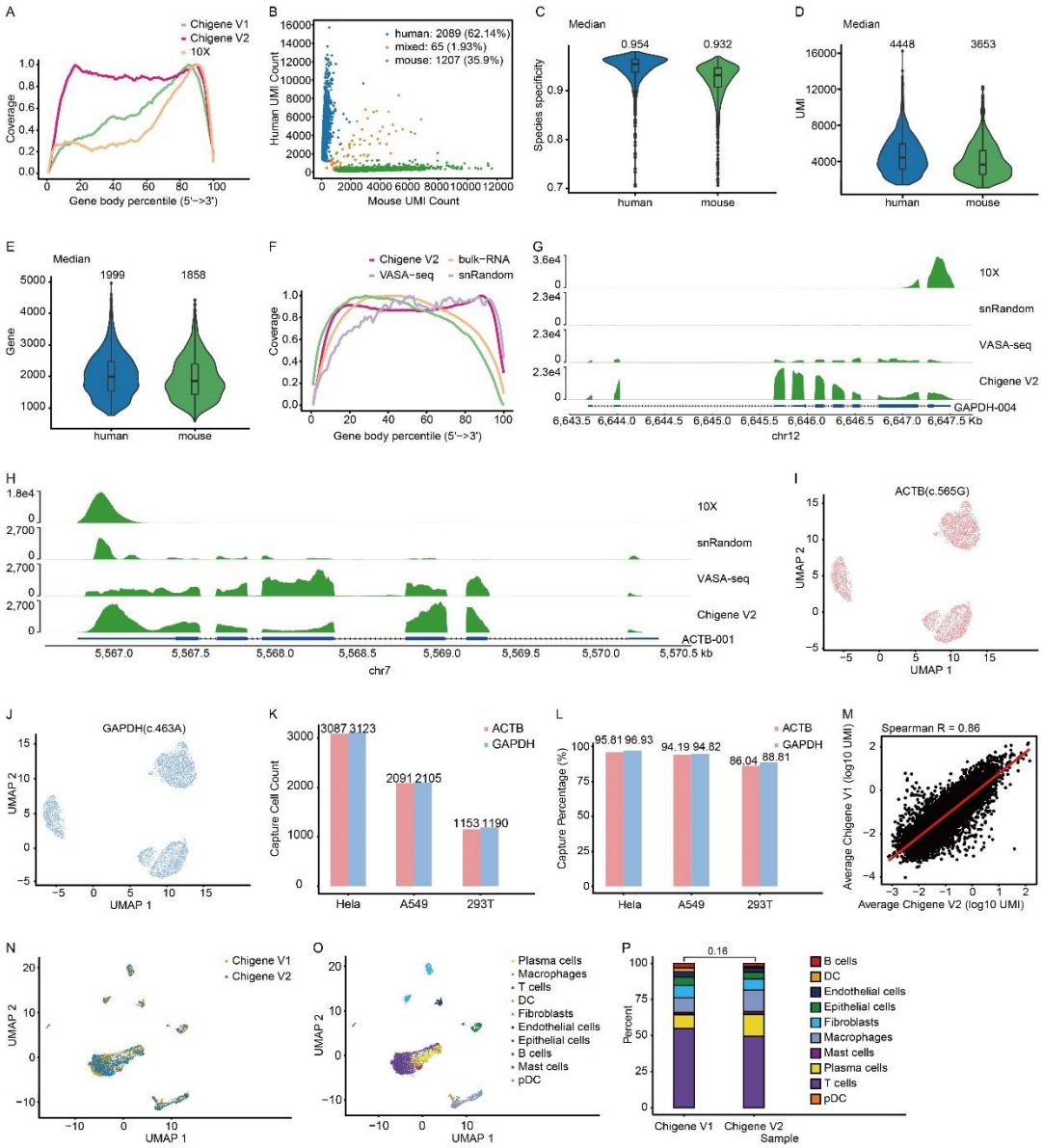


Figure 3. Development of the Chigene V2 platform based on the Chigene V1 platform and Random Bridging Co-labeling technique. (A) Gene body coverage shown by read distribution in three platforms. **(B)** Scatter plot illustrating the single-cell capture efficiency and doublet rate of Chigene V2-seq using species mixing. **(C)** Species specificity of unique molecular identifiers (UMIs) in an A549-MC38 cell mixture. Total number of identified A549 cells: n = 2089; total number of identified MC38 cells: n = 1207. The median species specificity of UMIs in A549 cells was 0.954, whereas the median species specificity of UMIs in MC38 cells was 0.932. **(D-E)** Violin plots and box plots depicting the number of UMIs and genes detected in individual A549

627 and MC38 cells. **(F)** Read distribution along the gene body compared between bulk RNA-seq and
628 three different full-length scRNA-seq methods (Chigene V2-seq, VASA, and snRandom RNA
629 profiling). **(G-H)** Representative raw reads aligned to the human genes GAPDH and ATCB via
630 four different scRNA-seq methods. **(I-J)** Clusters representing subpopulations in A549, 293T, and
631 HeLa cell mixtures in Chigene V2-seq, with the coverage of ATCB c.565G (red) and GAPDH
632 c.463A (blue) detected in each cluster. **(K-L)** Number and percentage of cells with coverage of
633 ATCB c.565G and GAPDH c.463A detected in each cluster. **(M)** Correlations between gene
634 expression measurements via the V2-seq and V1-seq methods. **(N)** UMAPs of single-cell profiles
635 (dots) from human urachal carcinoma libraries colored according to the different platforms used.
636 **(O)** UMAPs showing single-cell profiles (dots) from human urachal carcinoma libraries colored
637 by cell type. **(P)** Proportion of cells of each cell type (y-axis) detected via different methods (x-
638 axis).
639

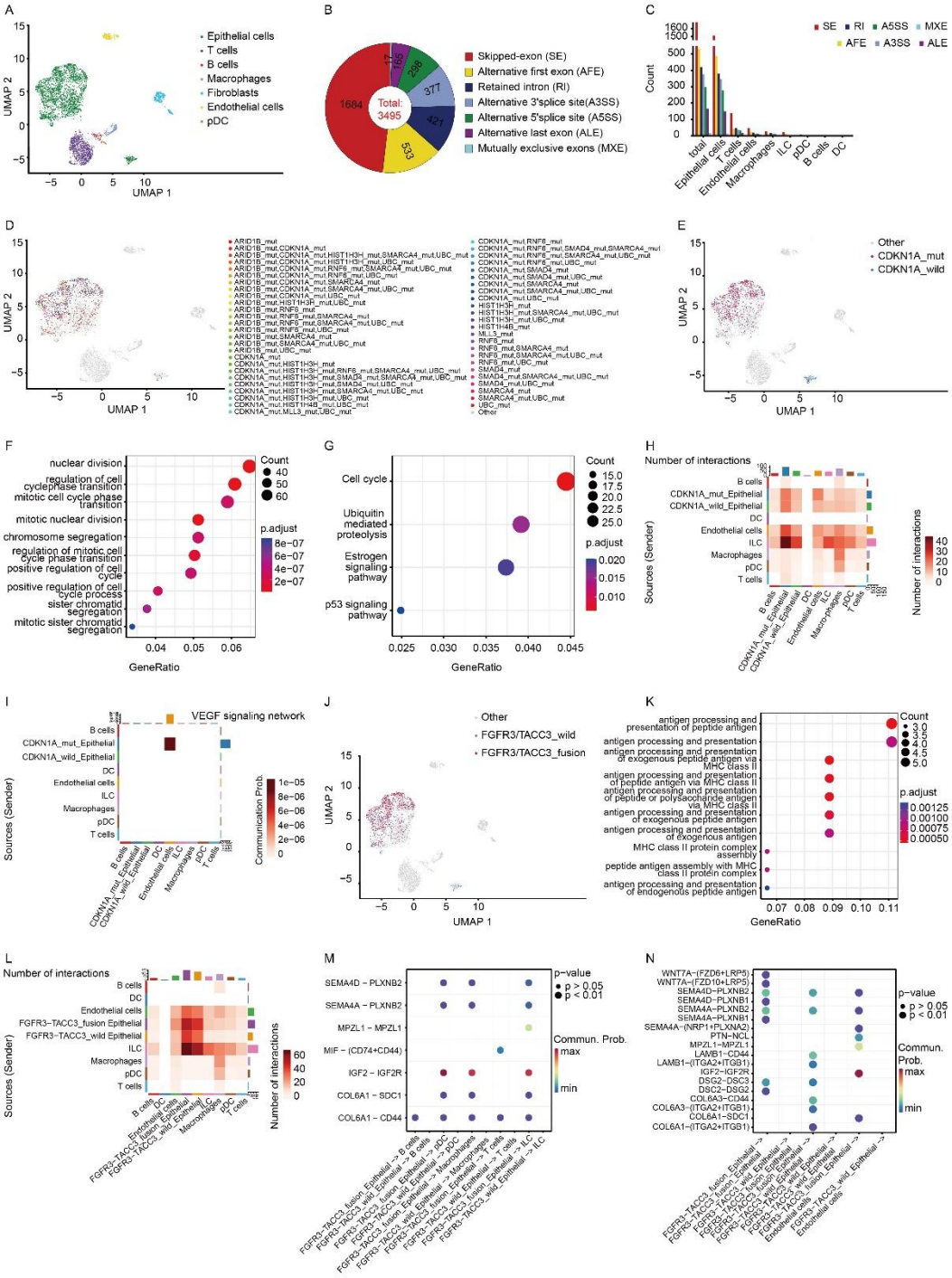


Figure 4. Identification of multi-landscape in clinical samples via the Chigene V2 platform. **(A)** UMAP representation of cells from the human urothelial cancer sample based on their gene expression, with seven annotated cell types displayed. **(B)** Pie plot illustrating the number of detected global gene AS events in Chigene V2-seq. **(C)** Abundance analysis of different classes of AS events across cell clusters. **(D)** Landscape visualization of mutated genes at the single-cell level in human urothelial cancer. **(E)** Detection of CDKN1A mutations in epithelial cells via

Chigene V2-seq. **(F-G)** GO and KEGG enrichment analyses of upregulated genes in CDKN1A-mutated cells. **(H)** Heatmap displaying the total number of interactions among different cell populations. **(I)** Heatmap showing the specific VEGF signaling network between CDKN1A-mutated epithelial and endothelial cells. **(J)** Detection of the FGFR3-TACC fusion gene in epithelial cells via Chigene V2-seq. **(K)** GO enrichment analysis of genes downregulated in FGFR3-TACC fusion cells. **(L)** Heatmap revealing the total number of interactions among different cell populations. **(M-N)** Bubble diagrams showing the probability of communication and statistical significance of receptor–ligand pairs between FGFR3-TACC fusion/wild-type epithelial and immune/nonimmune cells. The dot color represents the communication probability, reflecting the likelihood of successful information exchange between these cell types. The dot size represents the computed p value, which is used to assess the statistical significance of the observed interactions. Empty spaces indicate a communication probability of zero. P values were computed via a one-sided permutation test.

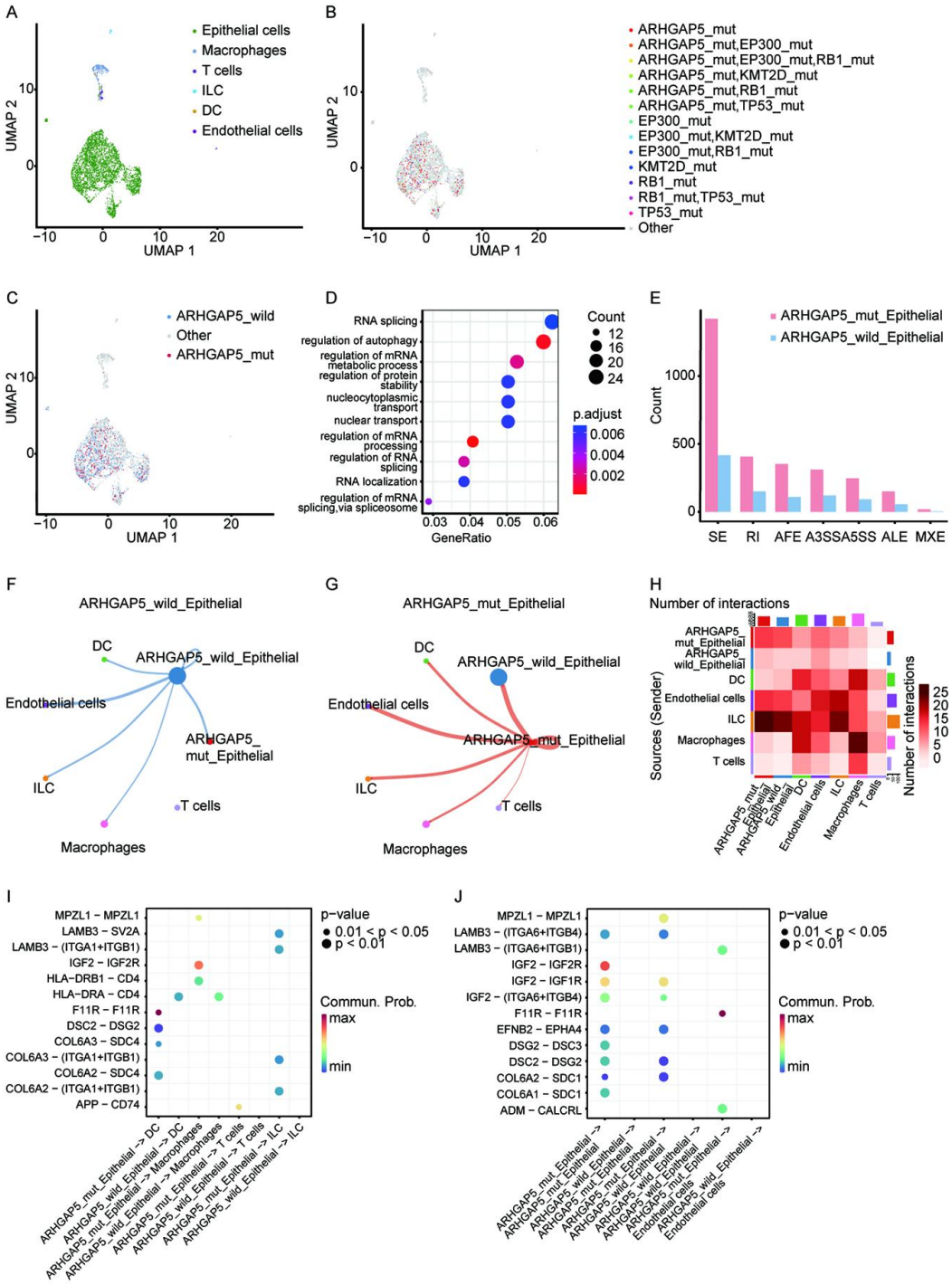


Figure 5. Chigene V2 enables the discovery of novel gene mutations in urothelial carcinoma at the single-cell level. (A) UMAP visualization of genes expressed in cells from a human urothelial cancer sample, with six annotated cell types displayed. **(B)** Landscape depiction of mutated genes at the single-cell level in human urothelial cancer. **(C)** Detection of the ARHGAP5 mutation in epithelial cells via Chigene V2-seq. **(D)** GO enrichment analysis of upregulated genes in ARHGAP5-mutated cells. **(E)** Comparison of AS events between ARHGAP5-mutated and wild-

type cells. **(F-G)** Total number of interactions between ARHGAP5-mutated/wild-type cells and other cell populations. The circle sizes represent the number of cells in each cell group, whereas the edge width represents the communication probability. **(H)** Heatmap displaying the total number of interactions among different cell populations. **(I-J)** Bubble diagrams showing the communication probability and statistical significance of receptor–ligand pairs between ARHGAP5-mutated/wild-type epithelial cells and immune/nonimmune cells. The dot color represents the communication probability, reflecting the likelihood of successful information exchange between these cell types. The dot size represents the computed p value, which is used to assess the statistical significance of the observed interactions. Empty spaces indicate a communication probability of zero. P values were computed via a one-sided permutation test.

References:

[1]. Haque, A., et al., A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. Genome Med, 2017. 9(1): p. 75.

[2]. Tanay, A. and A. Regev, Scaling single-cell genomics from phenomenology to mechanism. Nature, 2017. 541(7637): p. 331-338.

[3]. Wu, A.R., et al., Single-Cell Transcriptional Analysis. Annu Rev Anal Chem (Palo Alto Calif), 2017. 10(1): p. 439-462.

[4]. Hanahan, D. and R.A. Weinberg, Hallmarks of cancer: the next generation. Cell, 2011. 144(5): p. 646-74.

[5]. Bradley, R.K. and O. Anczukow, RNA splicing dysregulation and the hallmarks of cancer. Nat Rev Cancer, 2023. 23(3): p. 135-155.

[6]. Shi, X., et al., Mechanism insights and therapeutic intervention of tumor metastasis: latest developments and perspectives. Signal Transduct Target Ther, 2024. 9(1): p. 192.

[7]. Peng, Q., et al., Impacts and mechanisms of alternative mRNA splicing in cancer metabolism, immune response, and therapeutics. Mol Ther, 2022. 30(3): p. 1018-1035.

[8]. Chang, Y., Z. Zhao and Y. Song, Research progress on fusion genes in tumours. Clinical and Translational Discovery, 2024. 4(4).

[9]. Bonnal, S.C., I. Lopez-Oreja and J. Valcarcel, Roles and mechanisms of alternative splicing in cancer - implications for care. Nat Rev Clin Oncol, 2020. 17(8): p. 457-474.

[10]. Zeisel, A., et al., Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. Science, 2015. 347(6226): p. 1138-42.

[11]. Jaitin, D.A., et al., Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. Science, 2014. 343(6172): p. 776-9.

[12]. Grun, D., et al., Single-cell messenger RNA sequencing reveals rare intestinal cell types. Nature, 2015. 525(7568): p. 251-5.

[13]. Shalek, A.K., et al., Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. Nature, 2013. 498(7453): p. 236-40.

[14]. Wu, B., et al., Single-cell transcriptome analyses reveal critical roles of RNA splicing during leukemia progression. PLoS Biol, 2023. 21(5): p. e3002088.

[15]. Liu, F., et al., Systematic comparative analysis of single-nucleotide variant detection methods from single-cell RNA sequencing data. Genome Biol, 2019. 20(1): p. 242.

[16]. Zhang, T., et al., De novo identification of expressed cancer somatic mutations from single-cell RNA sequencing data. Genome Med, 2023. 15(1): p. 115.

[17]. Picelli, S., et al., Smart-seq2 for sensitive full-length transcriptome profiling in single cells. Nat Methods, 2013. 10(11): p. 1096-8.

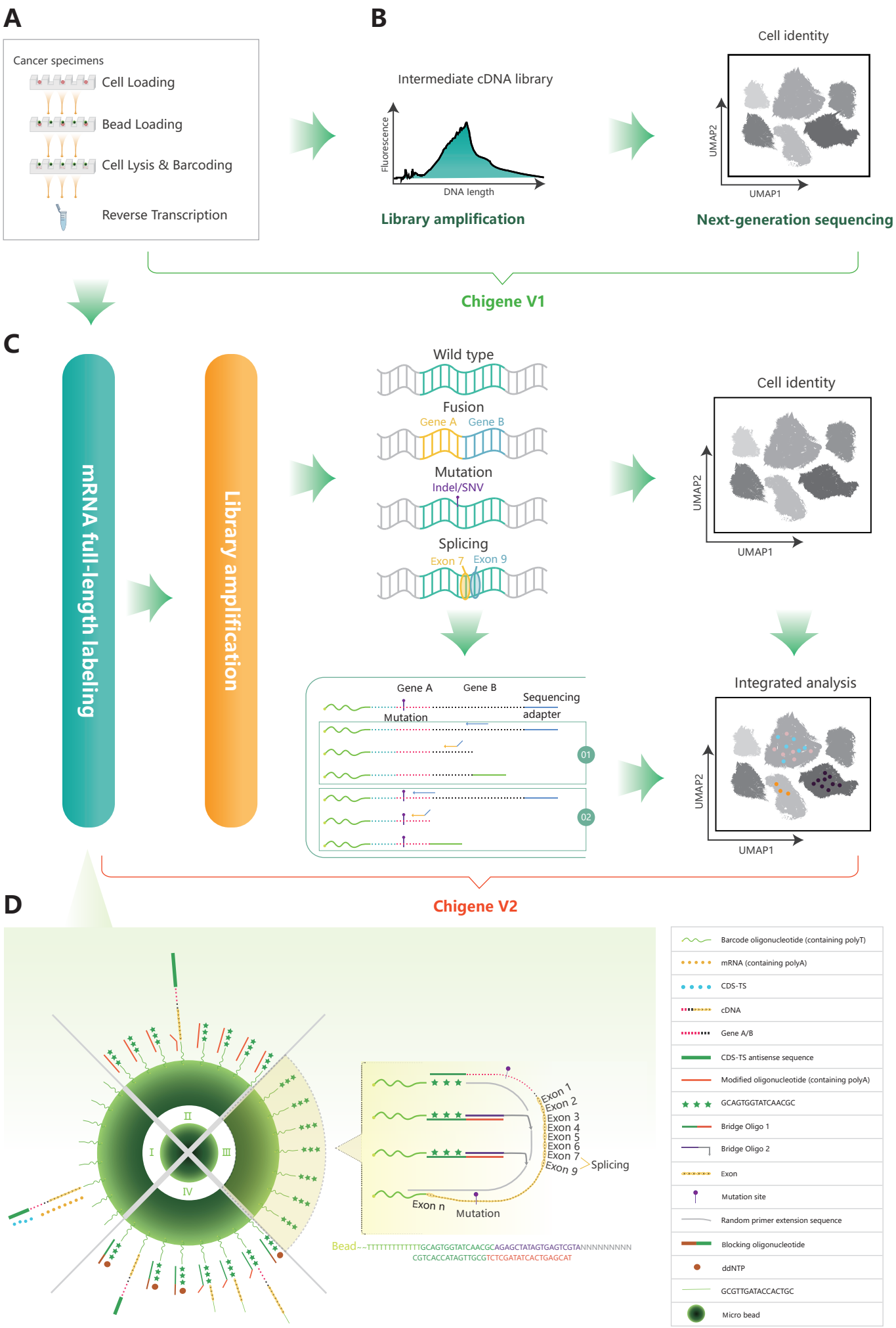
[18]. Xu, Z., et al., High-throughput single nucleus total RNA sequencing of formalin-fixed paraffin-embedded tissues by snRandom-seq. Nature communications, 2023. 14(1): p. 2734-2734.

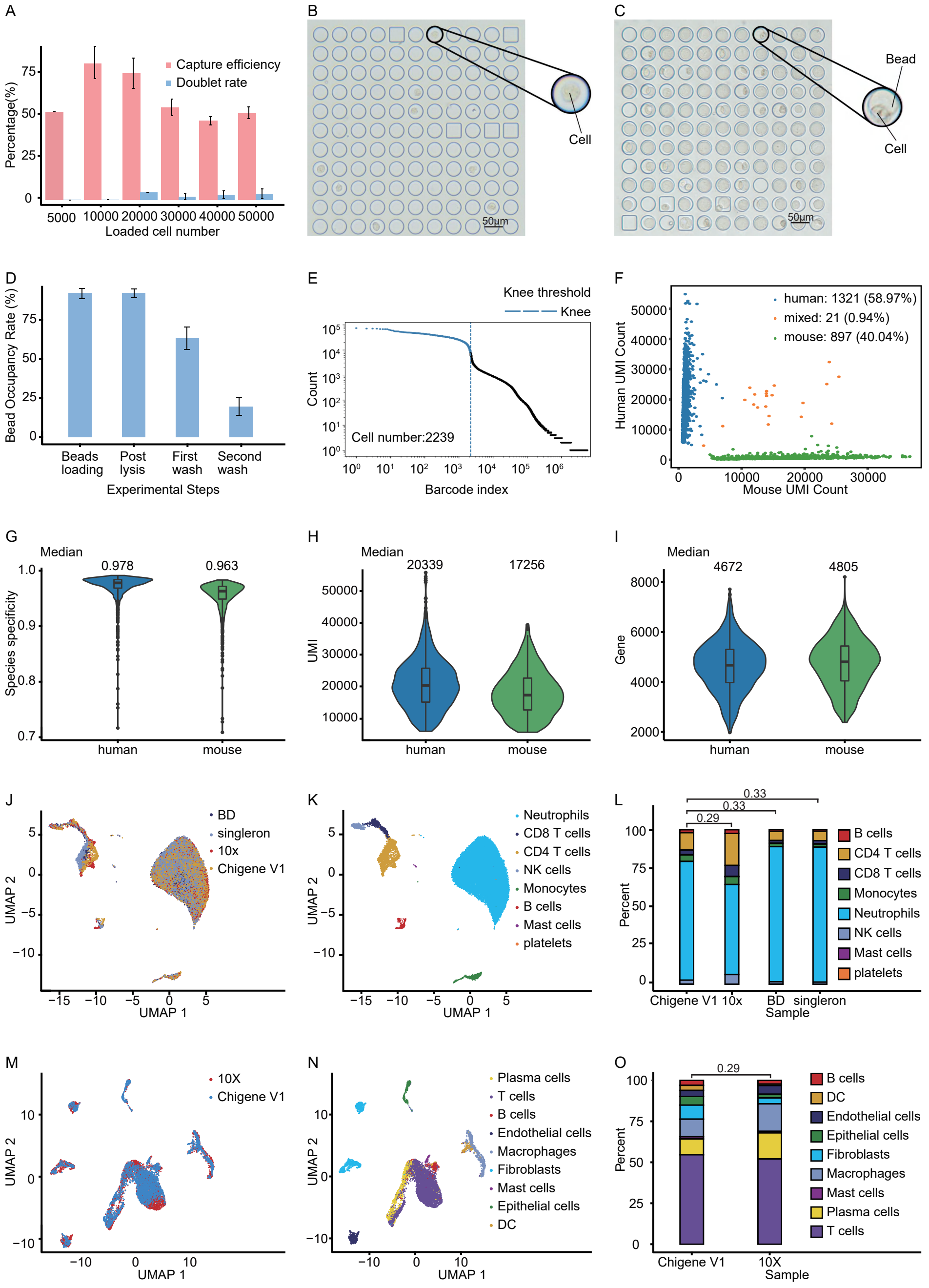
[19]. Salmen, F., et al., High-throughput total RNA sequencing in single cells using VASA-seq. Nat Biotechnol, 2022. 40(12): p. 1780-1793.

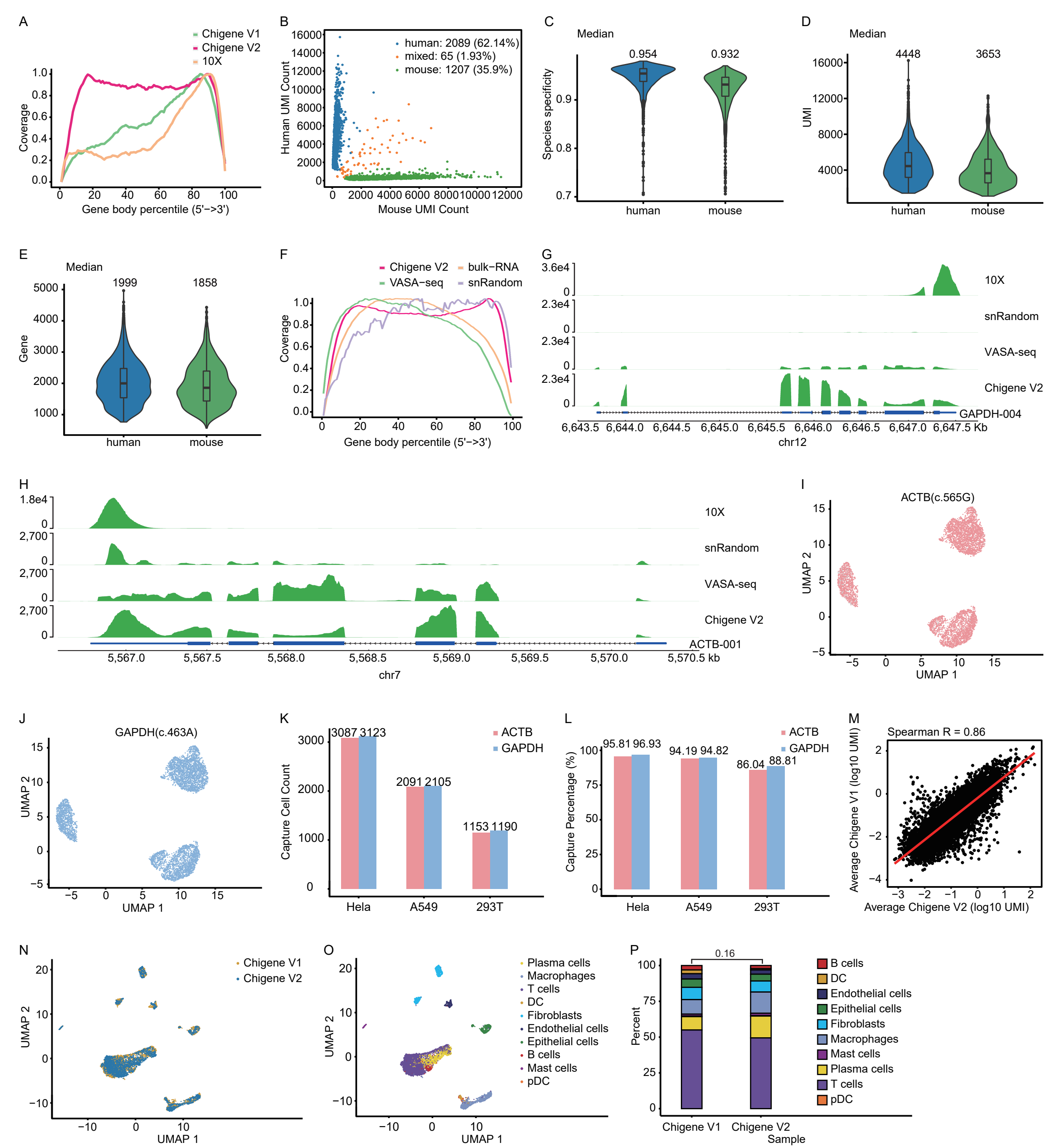
[20]. Chen, H., et al., Pan-Cancer Single-Nucleus Total RNA Sequencing Using snHH-Seq. Advanced Science, 2023.

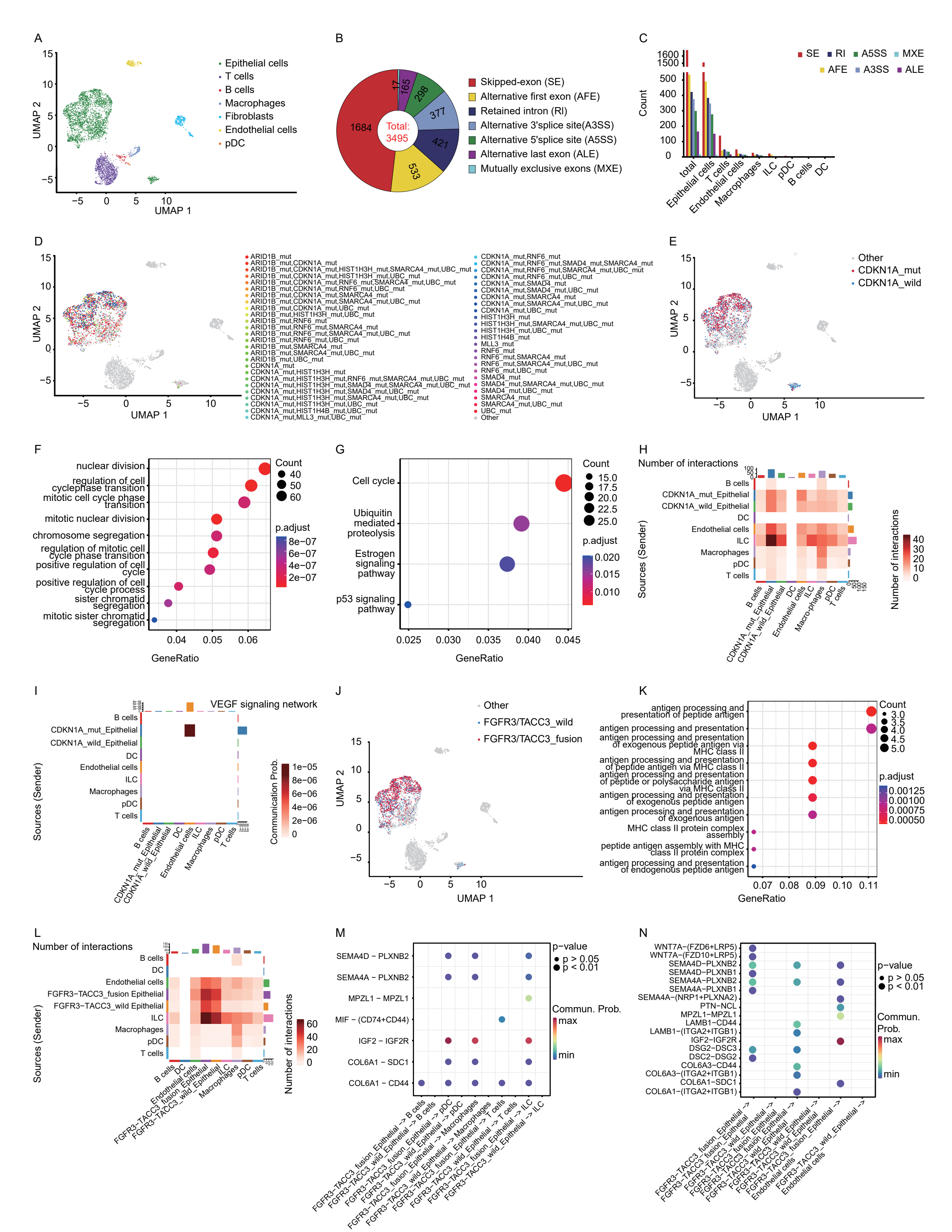
[21]. Jen, C.P., J.H. Hsiao and N.A. Maslov, Single-cell chemical lysis on microfluidic chips with arrays of microwells. Sensors (Basel), 2012. 12(1): p. 347-58.

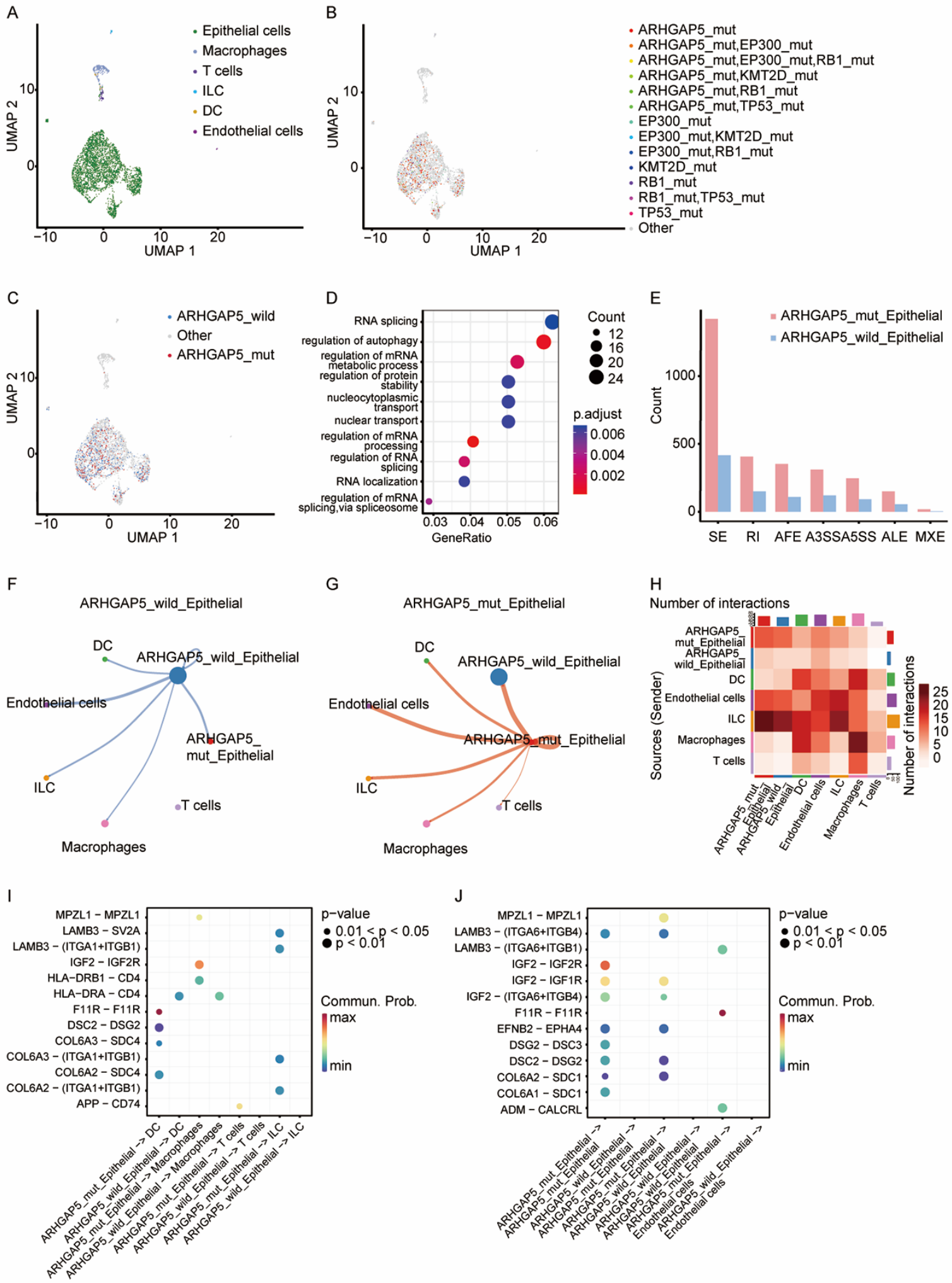
721 [22]. Gierahn, T.M., et al., Seq-Well: portable, low-cost RNA sequencing of single cells at high
722 throughput. Nat Methods, 2017. 14(4): p. 395-398.
723 [23]. Basu, A. and K. Kodumudi, Multimodal approaches to improve immunotherapy in breast cancer.
724 Immunotherapy, 2020. 12(3): p. 161-165.
725 [24]. Li, H., et al., Promotion of Sema4D expression by tumor-associated macrophages: Significance in
726 gastric carcinoma. World J Gastroenterol, 2018. 24(5): p. 593-601.
727 [25]. Chen, W.G., et al., Sema4D expression and secretion are increased by HIF-1alpha and inhibit
728 osteogenesis in bone metastases of lung cancer. Clin Exp Metastasis, 2019. 36(1): p. 39-56.
729 [26]. Takada, H., et al., Semaphorin 4D promotes bone invasion in head and neck squamous cell
730 carcinoma. Int J Oncol, 2017. 51(2): p. 625-632.
731 [27]. Jiang, H., et al., The role of semaphorin 4D in tumor development and angiogenesis in human
732 breast cancer. Onco Targets Ther, 2016. 9: p. 5737-5750.
733 [28]. Zhang, X., et al., High expression of COL6A1 predicts poor prognosis and response to
734 immunotherapy in bladder cancer. Cell Cycle, 2023. 22(5): p. 610-618.
735 [29]. Wang, A., et al., MPZL1 Promotes Lung Adenocarcinoma Progression by Enhancing Tumor
736 Proliferation, Invasion, Migration, and Suppressing Immune Function via Transforming Growth
737 Factor-beta1. Hum Gene Ther, 2023. 34(11-12): p. 540-553.
738 [30]. Chen, D., L. Cao and X. Wang, MPZL1 promotes tumor cell proliferation and migration via
739 activation of Src kinase in ovarian cancer. Oncol Rep, 2019. 42(2): p. 679-687.
740 [31]. Feng, J., et al., MPZL1 upregulation promotes tumor metastasis and correlates with unfavorable
741 prognosis in non-small cell lung cancer. Carcinogenesis, 2022. 43(10): p. 919-929.
742 [32]. Huang, X., et al., Wnt7a activates canonical Wnt signaling, promotes bladder cancer cell invasion,
743 and is suppressed by miR-370-3p. J Biol Chem, 2018. 293(18): p. 6693-6706.
744 [33]. Cera, M.R., et al., Increased DC trafficking to lymph nodes and contact hypersensitivity in
745 junctional adhesion molecule-A-deficient mice. J Clin Invest, 2004. 114(5): p. 729-38.
746 [34]. Leech, A.O., et al., Paradigms lost-an emerging role for over-expression of tight junction adhesion
747 proteins in cancer pathogenesis. Ann Transl Med, 2015. 3(13): p. 184.
748 [35]. Snyder, A., et al., Contribution of systemic and somatic factors to clinical response and resistance
749 to PD-L1 blockade in urothelial cancer: An exploratory multiomic analysis. PLoS Med, 2017. 14(5): p.
750 e1002309.
751 [36]. Braun, D.A., et al., Interplay of somatic alterations and immune infiltration modulates response to
752 PD-1 blockade in advanced clear cell renal cell carcinoma. Nat Med, 2020. 26(6): p. 909-918.
753 [37]. Tang, F., et al., mRNA-Seq whole-transcriptome analysis of a single cell. Nat Methods, 2009.
754 6(5): p. 377-82.
755 [38]. Nunes, L., et al., Prognostic genome and transcriptome signatures in colorectal cancers. Nature,
756 2024. 633(8028): p. 137-146.
757 [39]. Marzluff, W.F., E.J. Wagner and R.J. Duronio, Metabolism and regulation of canonical histone
758 mRNAs: life without a poly(A) tail. Nat Rev Genet, 2008. 9(11): p. 843-54.
759


A
Cancer specimens
Cell Loading
Bead Loading
Cell Lysis & Barcoding
Reverse Transcription
B
Intermediate cDNA library
Fluorescence
DNA length
Library amplification
Cell identity
UMAP2
UMAP1
Next-generation sequencing
Chigene V1
C
mRNA full-length labeling
Library amplification
Wild type
Fusion
Gene A
Gene B
Mutation
Indel/SNV
Splicing
Exon 7
Exon 9
Cell identity
UMAP2
UMAP1
Gene A
Gene B
Mutation
Sequencing adapter
01
02
Integrated analysis
UMAP2
UMAP1
Chigene V2
D
I
II
III
IV
Exon 1
Exon 2
Exon 3
Exon 4
Exon 5
Exon 6
Exon 7
Exon 9
Splicing
Exon n
Mutation
Bead~~TTTTTTTTTTTTTGCAGTGGTATCAACGCAGAGCTATAGTGAGTCGTANNNNNNNNN
CGTCACCATAGTTGCGTCTCGATATCACTGAGCAT
Barcode oligonucleotide (containing polyT)
mRNA (containing polyA)
CDS-TS
cDNA
Gene A/B
CDS-TS antisense sequence
Modified oligonucleotide (containing polyA)
GCAGTGGTATCAACGC
Bridge Oligo 1
Bridge Oligo 2
Exon
Mutation site
Random primer extension sequence
Blocking oligonucleotide
ddNTP
GCGTTGATACCACTGC
Micro bead


A
Capture efficiency
Doublet rate
Percentage(%)
Loaded cell number
5000 10000 20000 30000 40000 50000
B
Cell
50μm
C
Bead
Cell
50μm
D
Bead Occupancy Rate (%)
Beads loading
Post lysis
First wash
Second wash
Experimental Steps
E
Knee threshold
Knee
Count
Cell number:2239
Barcode index
F
human: 1321 (58.97%)
mixed: 21 (0.94%)
mouse: 897 (40.04%)
Human UMI Count
Mouse UMI Count
G
Median
0.978
0.963
Species specificity
human
mouse
H
Median
20339
17256
UMI
human
mouse
I
Median
4672
4805
Gene
human
mouse
J
BD
singleron
10x
Chigene V1
UMAP 2
UMAP 1
K
Neutrophils
CD8 T cells
CD4 T cells
NK cells
Monocytes
B cells
Mast cells
platelets
UMAP 2
UMAP 1
L
0.33
0.33
0.29
Percent
Chigene V1
10x
BD
singleron
Sample
B cells
CD4 T cells
CD8 T cells
Monocytes
Neutrophils
NK cells
Mast cells
platelets
M
10X
Chigene V1
UMAP 2
UMAP 1
N
Plasma cells
T cells
B cells
Endothelial cells
Macrophages
Fibroblasts
Mast cells
Epithelial cells
DC
UMAP 2
UMAP 1
O
0.29
Percent
Chigene V1
10X
Sample
B cells
DC
Endothelial cells
Epithelial cells
Fibroblasts
Macrophages
Mast cells
Plasma cells
T cells


A
Chigene V1
Chigene V2
10X
Coverage
Gene body percentile (5'->3')
B
human: 2089 (62.14%)
mixed: 65 (1.93%)
mouse: 1207 (35.9%)
Human UMI Count
Mouse UMI Count
C
Median
0.954
0.932
Species specificity
human
mouse
D
Median
4448
3653
UMI
human
mouse
E
Median
1999
1858
Gene
human
mouse
F
Chigene V2
bulk-RNA
VASA-seq
snRandom
Coverage
Gene body percentile (5'->3')
G
10X
snRandom
VASA-seq
Chigene V2
GAPDH-004
chr12
H
10X
snRandom
VASA-seq
Chigene V2
ACTB-001
chr7
I
ACTB(c.565G)
UMAP 2
UMAP 1
J
GAPDH(c.463A)
UMAP 2
UMAP 1
K
ACTB
GAPDH
3087 3123
2091 2105
1153 1190
Capture Cell Count
Hela
A549
293T
L
ACTB
GAPDH
95.81 96.93
94.19 94.82
86.04 88.81
Capture Percentage (%)
Hela
A549
293T
M
Spearman R = 0.86
Average Chigene V1 (log10 UMI)
Average Chigene V2 (log10 UMI)
N
Chigene V1
Chigene V2
UMAP 2
UMAP 1
O
Plasma cells
Macrophages
T cells
DC
Fibroblasts
Endothelial cells
Epithelial cells
B cells
Mast cells
pDC
UMAP 2
UMAP 1
P
0.16
Percent
Chigene V1
Chigene V2
Sample
B cells
DC
Endothelial cells
Epithelial cells
Fibroblasts
Macrophages
Mast cells
Plasma cells
T cells
pDC


A
Epithelial cells
T cells
B cells
Macrophages
Fibroblasts
Endothelial cells
pDC
UMAP 1
UMAP 2
B
Skipped-exon (SE)
Alternative first exon (AFE)
Retained intron (RI)
Alternative 3'splice site(A3SS)
Alternative 5'splice site (A5SS)
Alternative last exon (ALE)
Mutually exclusive exons (MXE)
Total: 3495
1684
533
421
377
298
165
17
C
SE
RI
A5SS
MXE
AFE
A3SS
ALE
Count
total
Epithelial cells
T cells
Endothelial cells
Macrophages
ILC
pDC
B cells
DC
D
E
Other
CDKN1A_mut
CDKN1A_wild
F
nuclear division
regulation of cell cyclephase transition
mitotic cell cycle phase transition
mitotic nuclear division
chromosome segregation
regulation of mitotic cell cycle phase transition
positive regulation of cell cycle
positive regulation of cell cycle process
sister chromatid segregation
mitotic sister chromatid segregation
Count
p.adjust
GeneRatio
G
Cell cycle
Ubiquitin mediated proteolysis
Estrogen signaling pathway
p53 signaling pathway
Count
p.adjust
GeneRatio
H
Number of interactions
B cells
CDKN1A_mut_Epithelial
CDKN1A_wild_Epithelial
DC
Endothelial cells
ILC
Macrophages
pDC
T cells
Sources (Sender)
Number of interactions
I
VEGF signaling network
Sources (Sender)
Communication Prob.
J
Other
FGFR3/TACC3_wild
FGFR3/TACC3_fusion
UMAP 1
UMAP 2
K
antigen processing and presentation of peptide antigen
antigen processing and presentation
antigen processing and presentation of exogenous peptide antigen via MHC class II
antigen processing and presentation of peptide antigen via MHC class II
antigen processing and presentation of peptide or polysaccharide antigen via MHC class II
antigen processing and presentation of exogenous peptide antigen
antigen processing and presentation of exogenous antigen
MHC class II protein complex assembly
peptide antigen assembly with MHC class II protein complex
antigen processing and presentation of endogenous peptide antigen
Count
p.adjust
GeneRatio
L
Number of interactions
Sources (Sender)
Number of interactions
M
SEMA4D − PLXNB2
SEMA4A − PLXNB2
MPZL1 − MPZL1
MIF − (CD74+CD44)
IGF2 − IGF2R
COL6A1 − SDC1
COL6A1 − CD44
p−value
Commun. Prob.
N
WNT7A−(FZD6+LRP5)
WNT7A−(FZD10+LRP5)
SEMA4D−PLXNB2
SEMA4D−PLXNB1
SEMA4A−PLXNB2
SEMA4A−PLXNB1
SEMA4A−(NRP1+PLXNA2)
PTN−NCL
MPZL1−MPZL1
LAMB1−CD44
LAMB1−(ITGA2+ITGB1)
IGF2−IGF2R
DSG2−DSC3
DSC2−DSG2
COL6A3−CD44
COL6A3−(ITGA2+ITGB1)
COL6A1−SDC1
COL6A1−(ITGA2+ITGB1)
p−value
Commun. Prob.


A
Epithelial cells
Macrophages
T cells
ILC
DC
Endothelial cells
UMAP 2
UMAP 1
B
ARHGAP5_mut
ARHGAP5_mut,EP300_mut
ARHGAP5_mut,EP300_mut,RB1_mut
ARHGAP5_mut,KMT2D_mut
ARHGAP5_mut,RB1_mut
ARHGAP5_mut,TP53_mut
EP300_mut
EP300_mut,KMT2D_mut
EP300_mut,RB1_mut
KMT2D_mut
RB1_mut
RB1_mut,TP53_mut
TP53_mut
Other
C
ARHGAP5_wild
Other
ARHGAP5_mut
D
RNA splicing
regulation of autophagy
regulation of mRNA metabolic process
regulation of protein stability
nucleocytoplasmic transport
nuclear transport
regulation of mRNA processing
regulation of RNA splicing
RNA localization
regulation of mRNA splicing,via spliceosome
GeneRatio
Count
p.adjust
E
ARHGAP5_mut_Epithelial
ARHGAP5_wild_Epithelial
Count
SE
RI
AFE
A3SS
A5SS
ALE
MXE
F
ARHGAP5_wild_Epithelial
DC
ARHGAP5_wild_Epithelial
Endothelial cells
ARHGAP5_mut_Epithelial
ILC
T cells
Macrophages
G
ARHGAP5_mut_Epithelial
DC
ARHGAP5_wild_Epithelial
Endothelial cells
ARHGAP5_mut_Epithelial
ILC
T cells
Macrophages
H
Number of interactions
Sources (Sender)
Number of interactions
I
MPZL1 − MPZL1
LAMB3 − SV2A
LAMB3 − (ITGA1+ITGB1)
IGF2 − IGF2R
HLA−DRB1 − CD4
HLA−DRA − CD4
F11R − F11R
DSC2 − DSG2
COL6A3 − SDC4
COL6A3 − (ITGA1+ITGB1)
COL6A2 − SDC4
COL6A2 − (ITGA1+ITGB1)
APP − CD74
p−value
0.01 < p < 0.05
p < 0.01
Commun. Prob.
max
min
J
MPZL1 − MPZL1
LAMB3 − (ITGA6+ITGB4)
LAMB3 − (ITGA6+ITGB1)
IGF2 − IGF2R
IGF2 − IGF1R
IGF2 − (ITGA6+ITGB4)
F11R − F11R
EFNB2 − EPHA4
DSG2 − DSC3
DSC2 − DSG2
COL6A2 − SDC1
COL6A1 − SDC1
ADM − CALCRL
p−value
0.01 < p < 0.05
p < 0.01
Commun. Prob.
max
min