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

Wenlong Zhong[1,3#], Ligang Wang[2#], Tonglei Guo[2], Lihua Zhao[2], Daqin Wu[1,3], Fei Xie[2], Xiao Wang[2], Xiuxin Li[2], Fangxiao Wang[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 multi-landscape of RNA. Here, we designed a microfluidic chip combined 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 using 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.

# 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 RNA alternative splicing[5] are all 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. 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[6-9], 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[10-13]. Recently, several high-throughput scRNA-seq platforms for total transcriptome full-length sequencing have emerged [16, 31, 33]. However, these methods pose limitations in terms of sequencing depth and mRNA detection sensitivity when targeting total RNA within a fixed amount of data.

To address the aforementioned 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. In 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 using 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 detection efficiency in single cell. Furthermore, we developed a simplified and cost-effective 3' sequencing approach by utilizing 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 overcomes the limitations of current scRNA-seq technologies, enabling comprehensive transcriptomic analysis at the single-cell level and expanding alternative 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 enabled the amplification and fragmentation labelling of mRNA using microbeads carrying oligo(dT) primers and random bridging primers on a microfluidic platform. This approach allows for 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). In Chigene V2, to further enhance the detection sensitivity of gene mutations and fusions in single cells, we employed a targeted mutation library enrichment strategy using specific primers. In brief, the process began with the capture of single-cell mRNA onto microbeads carrying oligo(dT) primers through reverse transcription on the microfluidic platform. Subsequently, the excess oligo(dT) on the microbeads was modified, followed by annealing, ligation, and extension using primer dimers carrying random sequences that were complementary to the modified oligo(dT). This step completed the random bridging step. The bridged and extended products were then amplified to generate sufficient cDNA for library construction and subsequent next-generation sequencing. To integrate the single-cell mRNA full-length gene expression profiles with the genomic features of interest, we called the molecular features of interest based on genomic features detected by whole exome sequencing (WES) from the same sample. Specifically, we enriched the library for the mutant and fusion genes of interest from the transcriptome library using PCR with specific primers. By extracting cell barcode and transcript information from the sequencing data, we improved the

detection rate of lowly expressed gene mutations. This allowed us to explore the biological functions of cell subpopulations with specific genomic features and their interactions with surrounding cells.

**Development and validation of the Chigene V1 platform based on a microfluidic chip**

To evaluate the performance of the microfluidic chip for single-cell sequencing, the Chigene V1 platform, which relies solely on chips and poly T primers, was first developed. To assess the cell capture efficiency and doublet rate of the microfluidic chip, different quantities of HeLa cells ranging from 5000 to 50000 cells were loaded. Microscopic counting revealed that the cell capture efficiency was generally greater than 50%, and the doublet rate was mostly less than 5% (Figure 2A-B). Observations of the microbeads after loading showed micropores with both single cells and beads (Figure 2C). The interval between micropore contamination was approximately 5 minutes, whereas cell lysis within the chip was achieved in approximately 12 seconds[14]. After the microbeads were loaded, cell lysis, RNA capture, and two quick washes were performed within two minutes to minimize cross-contamination. The recovery of a large portion of the microbeads (~70%) was achieved (Figure 2D). To evaluate the fidelity of the Chigene V1 platform, a standard mixed species experiment was conducted using cultured human (A549) and mouse (MC38) cell lines. A mixture of 2500 A549 cells and 2500 MC38 cells was profiled using Chigene V1-seq. Analysis identified 2239 high-quality unique cell barcodes, indicating clear separation of true cells from background noise (Figure 2E). The doublet rate was found to be low at 0.94% (Figure 2F), which is lower than that observed for other scRNA-seq methods (1.6-3.1%), such as Seq-Well[15] and VASAdrop[16]. Additionally, a high species specificity of UMI (~97%) was observed, indicating that high-fidelity single-cell libraries were produced by the Chigene V1 platform (Figure 2G). UMI and gene count distributions showed that the Chigene V1 platform captured a median of 20,339 UMIs and 4672 genes in single A549 cells, with an average of ~38k reads, and 17256 UMIs and 4805 genes in single MC38 cells, with an average of ~30k reads (Figure 2H-I). These results demonstrated the high sensitivity of the Chigene V1 platform.

Next, we evaluated the ability of the Chigene V1 platform to distinguish cell populations in complex primary samples. This was achieved by comparing its performance with that of 10x Genomics, BD Rhapsody, and Singleron using human peripheral blood cells following erythrocyte lysis. Compared with these mainstream scRNA platforms, Chigene V1-seq exhibited very high

124 cluster consistency (Figure 2J and Figure S1A-1D) and similar cell proportions (Figure 2K-L and
125 Figure S1E). However, the performance of the Chigene V1 platform was more similar to that of the
126 BD Rhapsody and Singleron platforms due to the use of the same cell capture strategy (microfluidic
127 chip). Furthermore, when a more complex urachal carcinoma specimen was used for performance
128 comparison between Chigene V1 and 10x Genomics, similar results were observed (Figure 2M-O,
129 Figure S1F-H).
130
131 **Development of the Chigene V2 platform based on the Chigene V1 platform and Random**
132 **Bridging Co-labeling technique**
133 To enhance the capture of full-length mRNAs, we integrated the Random Bridging Co-
134 labelling (RBCL) technique into the Chigene V1 platform and developed the Chigene V2
135 platform. The gene body coverage of the Chigene V2 platform was substantially greater than that
136 of the Chigene V1 and 10x platforms (Figure 3A). Moreover, species-mixing experiments were
137 conducted to analyse specificity on the Chigene V2 platform. The results revealed acceptable
138 doublet rates (1.93%) and specificity (~95%) in comparison to those of the Chigene V1 platform
139 (Figure 3B-C). Due to its ability to capture full-length genes, compared with the Chigene V1
140 platform, the Chigene V2 platform exhibited a decrease in average UMIs and gene counts per cell
141 at comparable sequencing depths (~36k reads per human cell and ~26 reads per mouse cell)
142 (Figure 3D-E). Additionally, the Chigene V2 platform demonstrated more homogeneous coverage
143 of gene body regions than did other high-throughput full-length single-cell RNA sequencing
144 platforms, such as VASA-seq, snRandom, and bulk RNA-seq (Figure 3F). Gene body coverage
145 analysis of the GAPDH and ACTB genes revealed that unlike poly(A)-based 10x Genomics,
146 which exhibited pronounced bias toward the 3'-end, Chigene V2, VASA-seq, and snRandom did
147 not display significant 3'- or 5'-end bias across the gene body. As expected, Chigene V2 displayed
148 a greater coverage depth, attributable to its emphasis on capturing mRNA within a comparable
149 sequencing data volume (Figure 3G-H).
150 The impressive ability of the Chigene V2 platform to detect full-length mRNA suggests that
151 it may be an ideal method for detecting mutations, gene fusions and alternative splicing at the
152 single-cell level. To improve the sensitivity of detecting mutations, particularly in low-abundance
153 transcripts and rare cells, a nested PCR method was employed for target enrichment of gene

154 mutations. To test this approach, a mixture of 293T, A549, and HeLa cell lines in equal
155 proportions was analysed using Chigene V2-seq to detect SNP sites in two housekeeping genes
156 (ACTB c.565G and GAPDH c.463A). The findings revealed the accurate identification of distinct
157 cell clusters representing the three cell lines, suggesting minimal cross-contamination associated
158 with the Chigene V2 platform (Figure 3I-J). Moreover, the detection of these two SNP sites in the
159 majority of cells underscores the strength of the Chigene V2 platform in conjunction with nested
160 PCR for mutation detection at the single-cell level. (Figure 3K-L). To evaluate whether the
161 Chigene V2 platform could provide sufficient information from clinical tissue samples similar to
162 the Chigene V1 platform, we collected fresh human urachal carcinoma samples and compared
163 their RNA profiles using Chigene V1-seq and Chigene V2-seq. The results showed a strong
164 correlation between the total RNA profiles obtained by Chigene V1-seq and Chigene V2-seq
165 (Spearman R: 0.86, Figure 3M). Furthermore, the majority of cell types were consistently
166 identified by both methods in the same sample (Figure 3N and Figure S2A-B). Unsupervised
167 clustering analysis identified ten types of cells in the merged sequencing data (Figure 3O and
168 Figure S2C). No significant differences were observed in the proportions of cell types between
169 Chigene V2-seq and Chigene V1-seq (Figure 3P).
170
171 **Identification of multi-landscape by Chigene V2 platform in the clinical samples**
172 To assess the performance of the Chigene V2 platform in exploring gene alternative splicing,
173 mutations, and fusions, fresh human urothelial cancer samples were subjected to Chigene V2-seq,
174 whole-exome sequencing (WES) and bulk RNA-seq. A total of 4452 cells were identified from the
175 single-cell suspension, with an average of 1696 genes and 4665 UMIs detected per cell. These
176 cells were clustered into seven distinct cell types (Figure 4A and Figure S3A). By profiling full-
177 length transcripts via the Chigene V2-seq platform, alternative splicing (AS) patterns across
178 different cell types were identified through the quantification of inclusion rates of non-overlapping
179 exonic regions. In this sample, seven different AS events (total: 3459 events) were detected
180 (Figure 4B), mainly in epithelial cells and T cells (Figure 4C). Notably, alternative first exon
181 (AFE) events were frequently detected (Figure 4B-C), further confirming the high coverage of
182 mRNA in the Chigene V2 platform. Using the combined WES and RNA-seq data, ten genes with
183 higher mutation frequencies at both the DNA and RNA levels were selected for targeted

enrichment in the Chigene V2 platform (Figure S3B). UMAP visualization revealed a multitude of gene mutations and co-mutations in epithelial cells (Figure 4D). To substantiate the reliability of the Chigene V2 platform and investigate the functions of the identified mutated genes at the single-cell level, particular attention was given to the CDKN1A p.F53Ffs*10 mutation (Figure 4E). Compared to those in CDKN1A wild-type epithelial cells, gene ontology (GO) and KEGG enrichment analyses revealed significant enrichment of upregulated genes related to cell cycle-related biological processes (Figure 4F) and signalling pathways (Figure 4G) in CDKN1A-mutated epithelial cells. Cell communication analysis revealed that compared with wild-type cells, CDKN1A-mutated epithelial cells had a greater number of interactions with endothelial cells (Figure 4H). Specifically, CDKN1A-mutated epithelial cells exhibited increased VEGF signalling to endothelial cells (Figure 4I), suggesting a potential role in promoting angiogenesis.

Next, we explored the fusion gene in the Chigene V2 platform and identified the fusion gene FGFR3-TACC in the single-cell dimension (Figure 4J and Figure S3C). GO enrichment analysis revealed that downregulated genes in FGFR3-TACC fusion epithelial cells were significantly enriched in antigen processing and presentation-related biological processes (Figure 4K), potentially indicating a contribution to immune evasion. Cell communication analysis demonstrated increased signal output from FGFR3-TACC fusion epithelial cells compared to that from wild-type FGFR3 epithelial cells (Figure 4L). Ligand–receptor pair analysis revealed that FGFR3-TACC fusion epithelial cells preferentially transmitted signals to both immune cells and non-immune cells via immunosuppressive ligand–receptor pairs and promoting angiogenesis and tumor proliferation ligand–receptor pairs (Figure 4M-N), such as SEMA4D-PLXNB1/2[17-21], COL6A1-SDC1/CD44[22], MPZL1-MPZL1[23-25], and WNT7A-(FZD6/10+LRP5) [26]. These findings suggested that the FGFR3-TACC fusion might hinder the efficacy of immunotherapy and contribute to poor prognosis by affecting both immune cells and non-immune cells. In summary, the Chigene V2 platform has the ability to explore alternative splicing, mutations, and fusions at the single-cell level.

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

To identify novel mutated genes, a fresh specimen of urothelial carcinoma was analysed using Chigene V2-seq. A total of 4491 cells with an average of 1821 genes and 4273 UMIs per

214 cell were identified. Unsupervised clustering revealed several distinct clusters representing
215 different cell types, including DCs, endothelial cells, epithelial cells, T cells, ILCs, and
216 macrophages (Figure 5A and Figure S4). Analysis of the Chigene V2-seq data revealed multiple
217 gene mutations in epithelial cells (Figure 5B). Our study specifically examined a mutated gene,
218 ARHGAP5, harboring the p.I307V mutation (Figure 5C). GO enrichment and alternative splicing
219 analyses demonstrated a greater frequency of RNA splicing events in ARHGAP5-mutated
220 epithelial cells (Figure 5D-E). Cell communication analysis further revealed that epithelial cells
221 with mutated ARHGAP5 displayed a distinct number of interactions with other clusters,
222 particularly notable in their interaction with T cells (Figure 5F-G). Subsequent examination of
223 ligand receptor pairs revealed notable differences between ARHGAP5-mutated and wild-type
224 tumor cells regarding the signals they transmitted to both immune and non-immune cells (Figure
225 5H). In contrast to ARHGAP5 wild-type epithelial cells, ARHGAP5-mutated epithelial cells
226 tended to transmit signals to dendritic cells (DCs) via the F11R-F11R interaction, which could
227 impede DC migration to lymph nodes and the initiation of tumor-specific immune responses [27].
228 Additionally, in ARHGAP5-mutated epithelial cells, signals were specifically transmitted to T
229 cells through APP-CD74, which is known to exert an inhibitory impact on the activation of CD4+
230 T cells (Figure 5I). Moreover, ARHGAP5-mutated epithelial cells also transmitted signals to non-
231 immune cells involved in tumorigenesis, metastasis, and angiogenesis (Figure 5J) via MPZL1-
232 MPZL1[23-25], F11R-F11R[28], and COL6A1-SDC1[22]. Consequently, it is speculated that
233 urothelial carcinoma patients with ARHGAP5 mutations may exhibit a poorer response to
234 immunotherapy.
235 To gain a more comprehensive understanding of the prevalence of ARHGAP5 mutations
236 across various cancers, we searched the TCGA database (Figure S5A). Intriguingly, our analysis
237 revealed a mutation frequency of approximately 2.6% for ARHGAP5 in urothelial carcinoma
238 (Figure S5B). To assess the clinical implications of these mutations, propensity score matching
239 (1:1) was conducted, matching patients with ARHGAP5 mutations to their wild-type counterparts
240 based on variables including cancer type, age, sex, race, and pathological stage (Figure S5C).
241 Despite not reaching statistical significance, an intriguing pattern emerged: progression-free
242 survival (PFS) tended to be shorter in the mutant patients (Figure S5D). The associations between
243 ARHGAP5 mutations and immunotherapy response in cohorts of patients who underwent

immunotherapy for urothelial carcinoma[29] or advanced clear cell renal cell carcinoma[30] 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 non-responders. 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. Chigene V1, a conventional single-cell sequencing platform utilizing 3' end sequencing, demonstrated equivalent performance and scalability compared to mainstream platforms such as 10X Genomics, BD Rhapsody, and Singleron. 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 primarily depend 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 [31-33]. However, the use of total RNA sequencing techniques presents inherent limitations in sequencing depth dedicated to mRNA due to the presence of non-coding 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 co-labeling technique, facilitating the specific capture of full-length mRNA. 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 mRNA within the same data volume.

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 due to 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, using the Chigene V2 platform, we not only confirmed the functionality of these genetic variations at the single-cell level as previously reported but also revealed their potential impact on surrounding cells at the cellular level. This has deepened and broadened our understanding of genes with known functions. Notably, using our newly developed single-cell platform, we conducted exploratory research focused on ARHGAP5 mutations and found that ARHGAP5 mutant tumor cells not only promoted tumor cell proliferation but also exhibited inhibitory effects on T cells. Further investigations into the impact of ARHGAP5 on progression-free survival (PFS) indicated that mutations in ARHGAP5 might be linked to shorter PFS and poorer treatment responses. Briefly, our platform provides a powerful tool for mapping the single-cell mutational landscape in human tumors, offering new insights and perspectives for

304 personalized treatment strategies and the discovery of novel targets.

305 Nevertheless, it is important to acknowledge certain limitations of our technique. First, the
306 single-cell cDNA amplification method relies on poly(T) primers, which restricts the detection of
307 mRNA molecules lacking a poly(A) tail, such as histone mRNAs[34]. Consequently, these specific
308 transcripts were not captured and detected by our mRNA-seq assay. Second, as our assay utilizes
309 mature mRNAs as templates, mutations present within intronic and intergenic regions are not
310 detected. These limitations could be addressed through future technical refinements.

311 In conclusion, the straightforward experimental protocols and comprehensive transcriptomic
312 information offered by our platform show promising potential for the significant contributions of
313 Chigene scRNA-seq to large-scale applications in both basic and clinical research in the future.

314

315 **Method**

316 **Microfluidic chip device fabrication and microbeads**

317 The chip comprised a support layer, chip microstructure layer, and chip top cover layer. The support
318 layer was composed of rigid glass, while all other layers were constructed using
319 polydimethylsiloxane (PDMS) material with mold casting. For the chip microstructure layer, a
320 photomask was created based on the chip design, and a mold was fabricated via traditional
321 photolithography methods (with an SU-8 negative photoresist). The resulting mold consisted of an
322 array of cylindrical microstructures with dimensions of 52 μm (height) and 38 μm (diameter) and a
323 center-to-center distance of 48 μm between cylinders. The PDMS mixture, consisting of a PDMS
324 base and curing agent at a ratio of 10:1, was adequately mixed and poured onto the chip
325 microstructure layer mold. After baking at 80°C for 30 minutes, the cured PDMS chip
326 microstructure layer was gently removed from the mold. Moving to the chip top cover layer, a mold
327 was prepared based on the chip design, including a chip chamber framework section and a top cover
328 section. Similarly, PDMS was cast onto the mold to form the chip top cover layer. To complete the
329 chip assembly, 1 mm diameter holes were punched in the top cover section, serving as the inlet and
330 outlet ports. The glass layer, PDMS chip microstructure layer, and PDMS chip top cover layer were
331 subsequently treated with plasma and bonded together to form the final microwell. The microbeads
332 were purchased from Chemgenes Corporation (MACOSKO-2011-10 B) and were filtered through
333 a 40 μm strainer.

334 **Species mixture experiment and human specimens**

335 HeLa cells, A549 cells, 293T cells, and MC38 cells were obtained from the CAMS Cell Culture
336 Center and cultured in Dulbecco's Modified Eagle's medium (DMEM, Gibco, Cat #11965092)
337 supplemented with 10% heat-inactivated fetal bovine serum (FBS, Gibco, Cat # 26010074). The
338 cells were maintained at 37 °C in a 5% $CO_2$ incubator (Thermo Heracell 240i) and passaged every
339 2 days. For the species mixture experiment, A549 and MC38 cells were harvested and washed three
340 times in PBS by centrifugation at 4 °C and 600 × g for 3 min. Subsequently, the cells were
341 individually counted, and equal numbers of cells from each cell line were mixed. The resulting cell
342 mixture was subjected to single-cell RNA sequencing (scRNA-seq) following the Chigene scRNA-
343 seq protocol. The collection of human peripheral blood and urological cancer samples, as well as
344 the research conducted in this study, was approved by the Research Ethics Committee of Sun Yat-
345 sen Memorial Hospital, Sun Yat-sen University (approval numbers: SYSKY-2023-396-01). Clinical
346 information was collected after written informed consent was obtained.

347 **Library preparation**

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

362 **Random bridging co-labeling for Chigene V2-seq**

363 To obtain full-length mRNA, after reverse transcription and template switching, the mRNA-

364 free barcode oligo was annealed with the tag primer using the following reaction mixture: 100,000
365 microbeads in 35 μl of NF-water, 5 μl of 10 μM tag primer, and 10 μl of 20X SSC Buffer (NEB).
366 The reaction mixture was incubated at room temperature for 15 minutes. After annealing, the
367 microbeads were washed with PBS. To modify the 3' end of the barcode oligo, an extension mixture
368 consisting of 2 μl of NEB Buffer 10X, 5 μl of dNTPs, 2 μl of Klenow (exo-), and 11 μl of NF-water
369 was added. The mixture was then incubated at 37°C for 30 minutes. Denature buffer (Illumina) was
370 added to remove the mRNA and tag primer from the microbeads. The microbeads were then washed
371 twice with PBS. Bridge oligo1 and bridge oligo2 were annealed to form a bridge dimer. One end of
372 the bridge dimer was reverse complemented to the modified barcode oligo, while the other end
373 contained tens of random nucleotides "N" at the 3' tail. When the modified barcode oligo annealed
374 with the bridge dimer, the 3' end of the barcode oligo joined the neighboring 5' end of the N-tail
375 hand. The bridge capture mixture, which consisted of 10 μl of 20X SSC buffer, 10 μl of 10 μM
376 bridge dimer, and 100,000 microbeads in 30 μl of NF-water, was incubated at room temperature for
377 15 minutes. The microbeads were then washed with PBS and suspended in the ligation reaction
378 mixture, which included 2 μl of 10X T4 DNA ligase buffer, 5 μl of ATP, 2 μl of T4 DNA ligase, and
379 11 μl of NF-water. This mixture was incubated at room temperature for 30 minutes. After washing
380 with PBS, the extension was performed using the following reaction mixture: 2 μl of NEB Buffer
381 10X, 5 μl of dNTPs, 2 μl of Klenow (exo-), and 11 μl of NF-water. To improve efficiency, it is
382 recommended to repeat the steps from bridge capture to extension 2-3 times. To block the free
383 barcode oligo, ddNTPs were used, and then the cDNA was amplified and purified.

384 **Targeted capture of specific regions of genes**

385 For the confirmation of mutation sites or gene fusions by WES, nested PCR primers were designed
386 upstream and downstream of each target region. The upstream-out primers, downstream-out primers,
387 upstream-in primers, and downstream-in primers were mixed separately. A total of 2.5 ng of cDNA
388 was used as a template for amplification with the upstream and downstream primers. Subsequently,
389 secondary PCRs were carried out using the upstream and downstream primers. After each round of
390 amplification, the reaction products were individually purified using magnetic beads. The purified
391 products from each reaction were then combined. Index PCR was performed on the combined
392 products to generate a sequencing library. All libraries were subjected to quality control analysis
393 using an Equalbit 1× dsDNA HS Assay Kit (Vazyme, China) 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 using 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 GRCm39 hybrid genome) using the STAR aligner (version 2.7.1a) with default parameters. Reads uniquely mapped to the reference genome were assigned to genes using 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 using 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 using these highly variable genes. A Shared Nearest Neighbor (SNN) graph was constructed using the top 15 principal components, and cells were clustered using 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 using celltypist (version 0.1.9).

**Expression correlation analysis for the Chigene V1/V2 platform.** To compare the gene expression correlation between the Chigene V1 and Chigene V2 platforms, we imported the expression matrices of both platforms into Seurat (version 3.0.2) and performed data normalization and standardization using the same default parameters. Then, we utilized the Seurat function "AverageExpression" to calculate the average expression levels of genes in each platform. Finally, we used the "cor" function in R (version 4.0.2) to calculate the Spearman correlation coefficient based on the natural logarithm of the average expression levels.

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

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

**Cell interaction analysis.** Cell communication analysis was performed using the R package CellChat (version 1.5.0) with default parameters.

**GO and KEGG analysis.** For the enrichment analysis of DEGs between comparison groups, the FindMarkers function in Seurat was used. This function allows the identification of differential gene sets based on specified criteria such as the percentage of cells expressing the gene (pct) and the p value (p_val). For example, genes with pct > 0.1 and p_val < 0.05 were selected. To perform Gene

Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses of the DEGs, the R package clusterProfiler was used. This package provides functions for functional enrichment analysis and visualization. By default, the package uses hypergeometric tests for enrichment analysis and includes a set of predefined gene sets from databases such as GO and KEGG.

**Reporting summary**

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

**Data availability**

The data supporting the findings of this study have been deposited in the Genome Sequence Archive (GSA) at the National Genomics Data Center, China National Center for Bioinformation/Beijing Institute of Genomics, Chinese Academy of Sciences. The data can be accessed at https://ngdc.cncb.ac.cn/gsa-human, with the reference number PRJCA023899. The deposited data are compliant with the regulations set by the Ministry of Science and Technology of China and have been made publicly available.

**Acknowledgements**

This study was supported by the National Key Research and Development Program of China (Grant No. 2022YFC2408300), the National Natural Science Foundation of China (Grant No. 82322056, 82341018, 82072827, 81825016, and 82273421), the Guangdong Basic and Applied Basic Research Foundation (Grant No. 2021B1515020009), the Science and Technology Program 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).

**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. Lihua Zhao performed the data analysis, 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 authors have read and approved the final version of the manuscript.

**Conflicts of interest**

LG Wang and WY 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. (WY Gu is the founder; LG Zang, Tonglei Guo, LH Zhao, Fei Xie, Xiao Wang, Xiuxin Li and Fangxiao Wang are employees). The remaining authors declare no competing interests.

499 **Figure legends**

500 **Figure 1. Schematic workflow of Chigene-seq based on the microfluidic chip platform. (A)**
501 Unique mRNA capture was performed using oligo(dT) on a microfluidic chip, and the recovered
502 microbeads were subjected to reverse transcription. **(B)** An intermediate cDNA library was
503 amplified and subjected to next-generation sequencing for conventional 3' scRNA-seq (Chigene
504 V1). **(C)** To obtain overall information of mRNA molecular features (Chigene V2), full-length
505 mRNA labelling was performed by modifying the barcode oligo and performing random bridging
506 ligation and extension. After amplifying the molecular feature library, one portion was used for
507 identifying cell clusters, while the other portion was used for target enrichment of interesting
508 molecular features. Finally, cell identity and molecular features were integrated to analyse specific
509 mutations, fusions, or alternative splicing events at the single-cell level.

510

511 **Figure 2. Development and validation of the Chigene V1 platform based on the microfluidic**
512 **chip. (A)** Capture rate and doublet rate at various cell loading numbers. **(B)** Bright-field images
513 depicting a region of the device after cell loading. **(C)** Bright-field images showing a cell/bead-
514 loaded device. **(D)** Bead occupancy rate following bead loading, cell lysis, and two rounds of
515 washing. **(E)** Barcode plot used for identifying barcodes representing true cells (blue line). The
516 barcodes of the A549-MC38 mixed cells were arranged in descending order based on gene counts.
517 **(F)** Species-mixing scatter plot showing the single cell capture efficiency and doublet rate of
518 Chigene V1-seq. **(G)** Species specificity of UMIs in human-mouse cell mixture. Median of species
519 specificity of UMIs in human cells(A549) was 0.978. Median of species specificity of UMIs in
520 mouse cells (MC38) was 0.963. **(H-I)** Violin plots and box plots showing the number of UMIs and
521 genes detected in individual A549 and MC38 cells. A549 cells: n = 1321, MC38 cells: n = 897. The
522 data depicted in the box plots correspond to the first and third quartiles (lower and upper hinges)
523 and the median (centre). **(J and K)** UMAPs displaying single-cell profiles (dots) from PBMCs with
524 colors indicating the different platforms used (J) and each cell type (K). **(L)** Proportion of cells of
525 each cell type (y-axis) detected using different platforms (x-axis) in the PBMC libraries. **(M and N)**
526 UMAPs displaying single-cell profiles (dots) of human urachal carcinoma cells with colors
527 indicating the different platforms used (M) and each cell type (N). **(O)** Proportion of cells of each
528 cell type (y-axis) detected using Chigene V1 and 10x Genomics (x-axis) in human urachal

529 carcinoma libraries.

530

531 **Figure 3. Development of the Chigene V2 platform based on the Chigene V1 platform and**
532 **Random Bridging Co-labeling technique. (A)** Gene body coverage shown by read distribution
533 in three platforms. **(B)** Scatter plot illustrating the single-cell capture efficiency and doublet rate of
534 Chigene V2-seq using species mixing. **(C)** Species specificity of unique molecular identifiers
535 (UMIs) in an A549-MC38 cell mixture. Total number of identified A549 cells: n = 2089; total
536 number of identified MC38 cells: n = 1207. The median species specificity of UMIs in A549 cells
537 was 0.954, while the median species specificity of UMIs in MC38 cells was 0.932. **(D-E)** Violin
538 plots and box plots depicting the number of UMIs and genes detected in individual A549 and
539 MC38 cells. **(F)** Read distribution along the gene body compared between bulk RNA-seq and
540 three different full-length scRNA-seq methods (Chigene V2-seq, VASA, and snRandom RNA
541 Profiling). **(G-H)** Representative raw reads aligned to the human genes GAPDH and ATCB via
542 four different scRNA-seq methods. **(I-J)** Clusters representing subpopulations in an A549, A293T,
543 and HeLa cell mixture in Chigene V2-seq, with the coverage of ATCB c.565G (red) and GAPDH
544 c.463A (blue) detected in each cluster. **(K-L)** Number and percentage of cells with coverage of
545 ATCB c.565G and GAPDH c.463A detected in each cluster. **(M)** Correlations between gene
546 expression measurements via the Chigene V2-seq and Chigene V1-seq methods. **(N)** UMAPs of
547 single-cell profiles (dots) from human urachal carcinoma libraries colored according to different
548 platforms used. **(O)** UMAPs showing single-cell profiles (dots) from human urachal carcinoma
549 libraries colored by cell type. **(P)** Proportion of cells of each cell type (y-axis) detected using
550 different methods (x-axis).

551

552 **Figure 4. Identification of multi-landscape in clinical samples via the Chigene V2 platform.**
553 **(A)** UMAP representation of cells from the human urothelial cancer sample based on their gene
554 expression, with seven annotated cell types displayed. **(B)** Pie plot illustrating the number of
555 detected global gene AS events in Chigene V2-seq. **(C)** Abundance analysis of different classes of
556 AS events across cell clusters. **(D)** Landscape visualization of mutated genes at the single-cell
557 level in human urothelial cancer. **(E)** Detection of CDKN1A mutations in epithelial cells by
558 Chigene V2-seq. **(F-G)** GO and KEGG enrichment analysis 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 by 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/non-immune cells. The dot color represents the communication probability, and the dot size represents the computed p value. Empty spaces indicate a communication probability of zero. P values were computed using 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 cells from the human urothelial cancer sample based on their gene expression, 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 by 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, while 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/non-immune cells. The dot color represents the communication probability, and the dot size represents the computed p value. Empty spaces indicate a communication probability of zero. P values were computed using 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]. 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.

[7]. 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.

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

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

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

[11]. 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.

[12]. 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.

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

[14]. 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.

[15]. Gierahn, T.M., et al., Seq-Well: portable, low-cost RNA sequencing of single cells at high throughput. Nat Methods, 2017. 14(4): p. 395-398.

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

[17]. Basu, A. and K. Kodumudi, Multimodal approaches to improve immunotherapy in breast cancer. Immunotherapy, 2020. 12(3): p. 161-165.

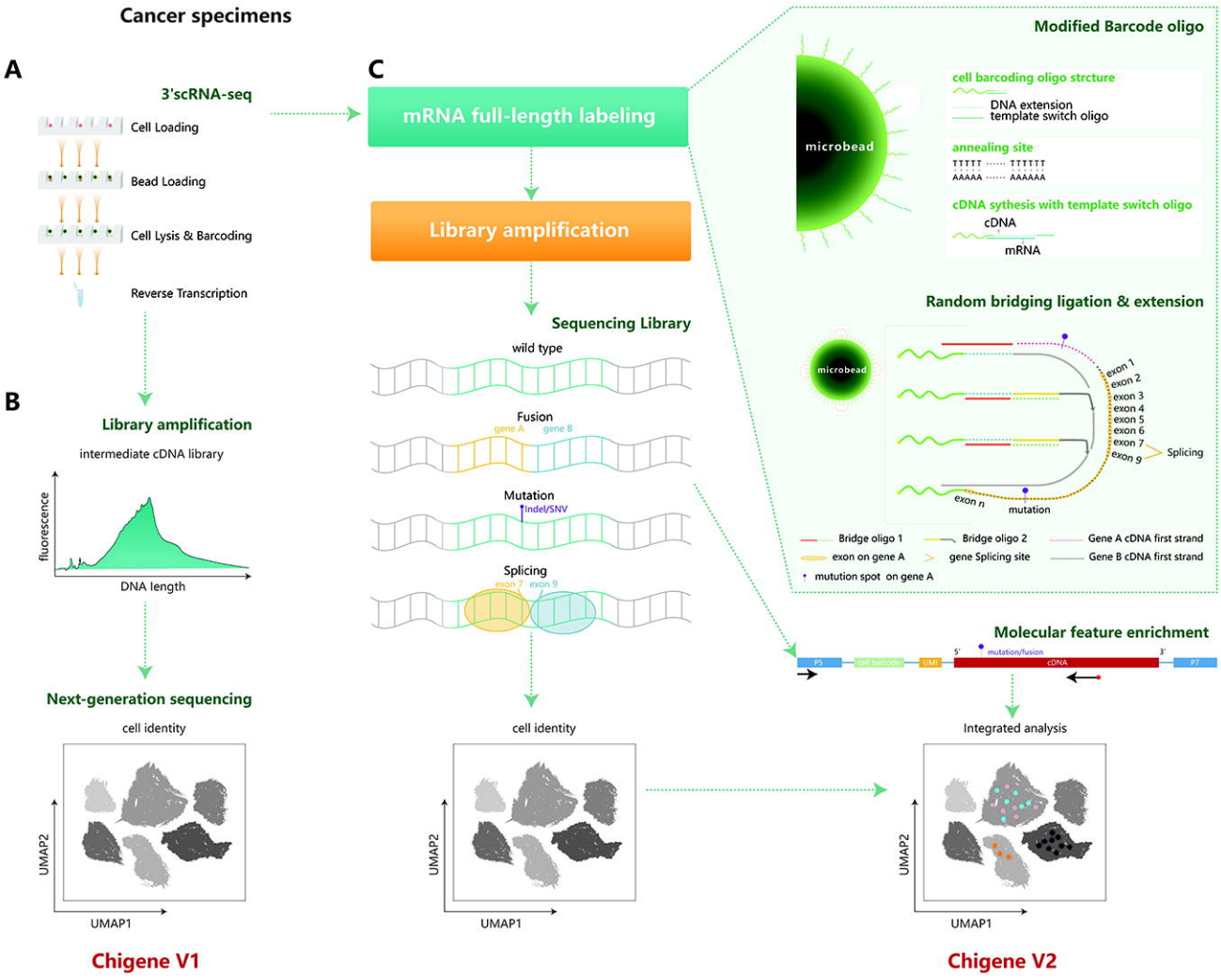
[18]. Li, H., et al., Promotion of Sema4D expression by tumor-associated macrophages: Significance in gastric carcinoma. World J Gastroenterol, 2018. 24(5): p. 593-601.

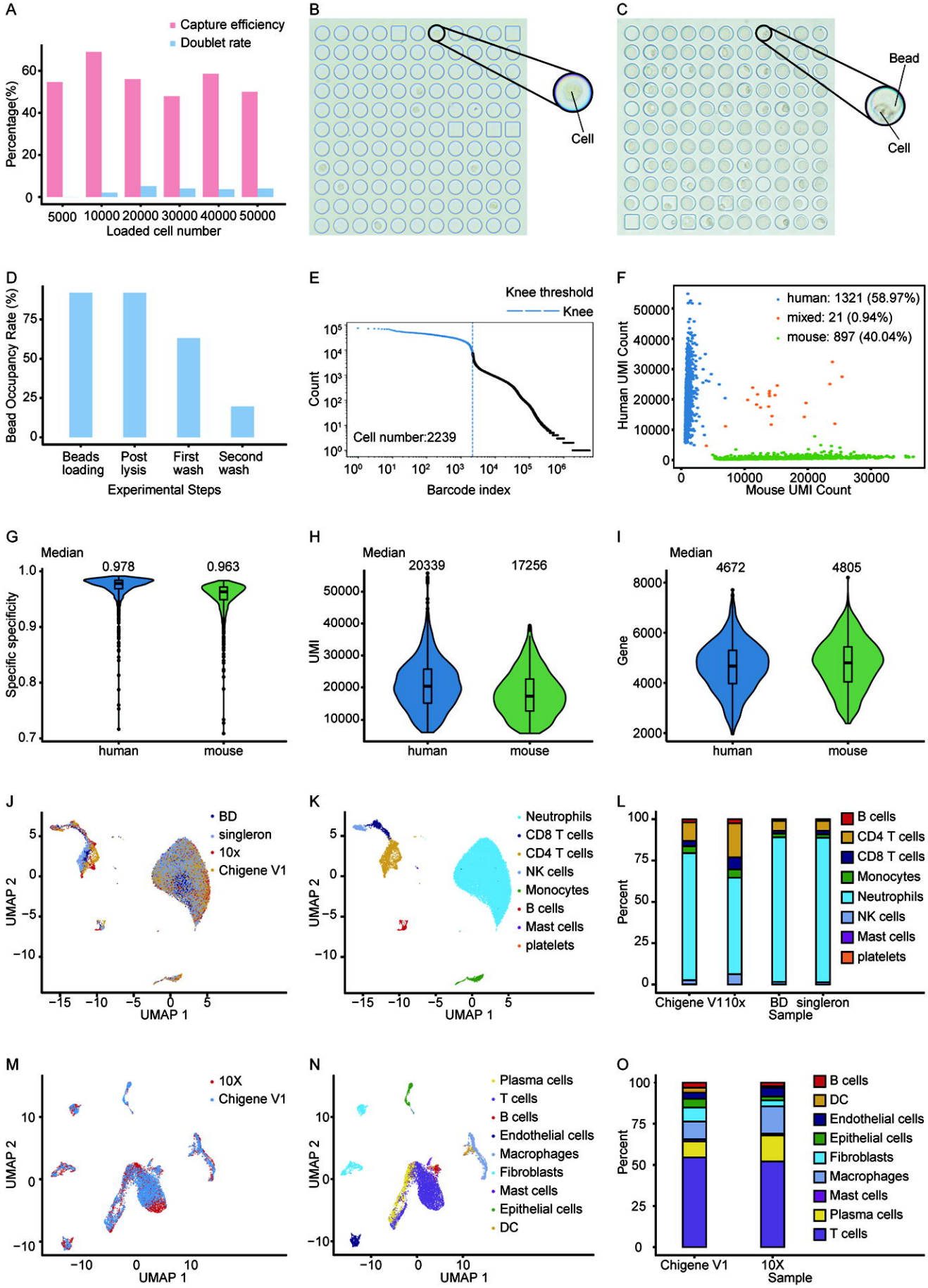
[19]. Chen, W.G., et al., Sema4D expression and secretion are increased by HIF-1alpha and inhibit osteogenesis in bone metastases of lung cancer. Clin Exp Metastasis, 2019. 36(1): p. 39-56.

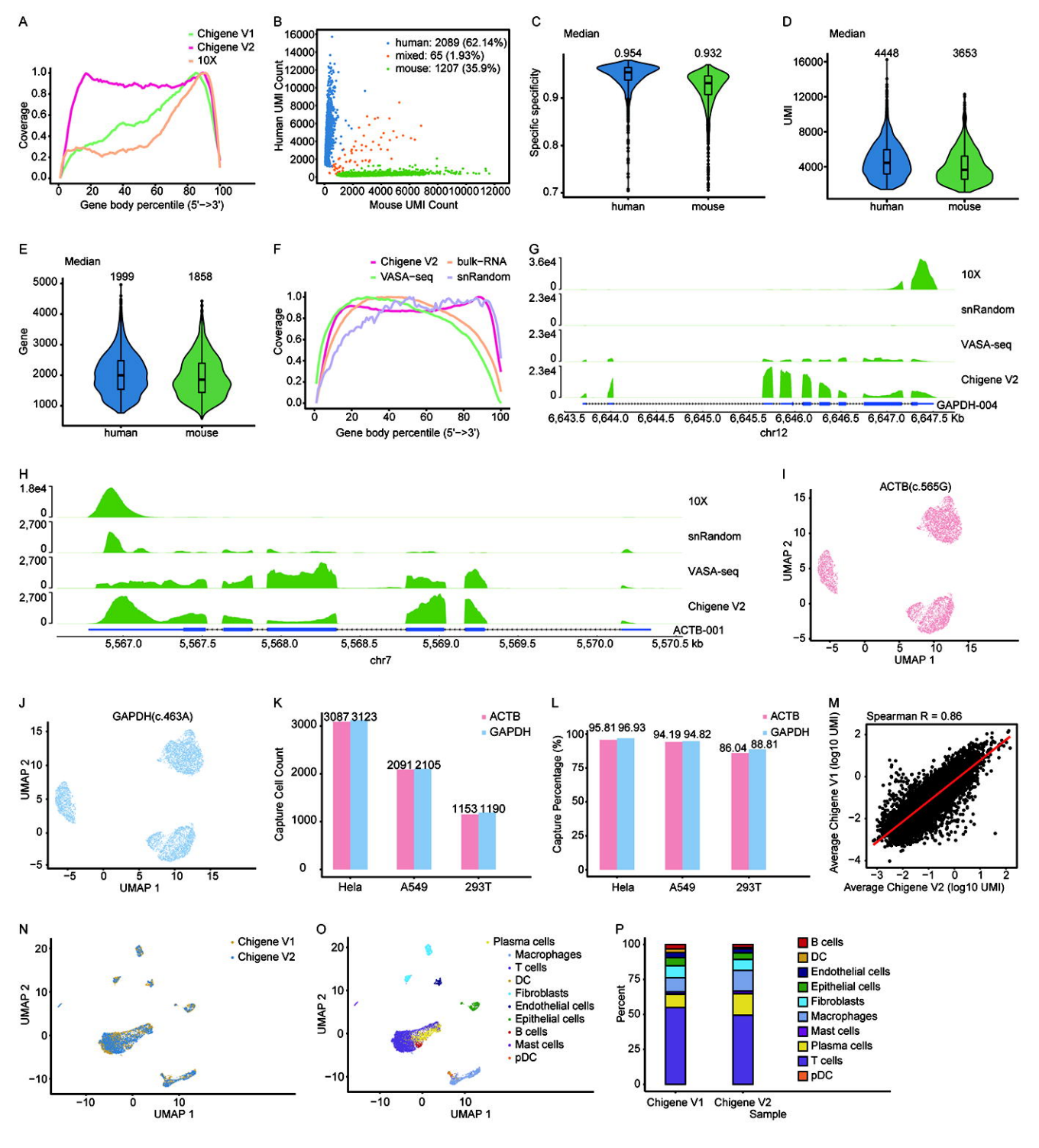
[20]. Takada, H., et al., Semaphorin 4D promotes bone invasion in head and neck squamous cell carcinoma. Int J Oncol, 2017. 51(2): p. 625-632.

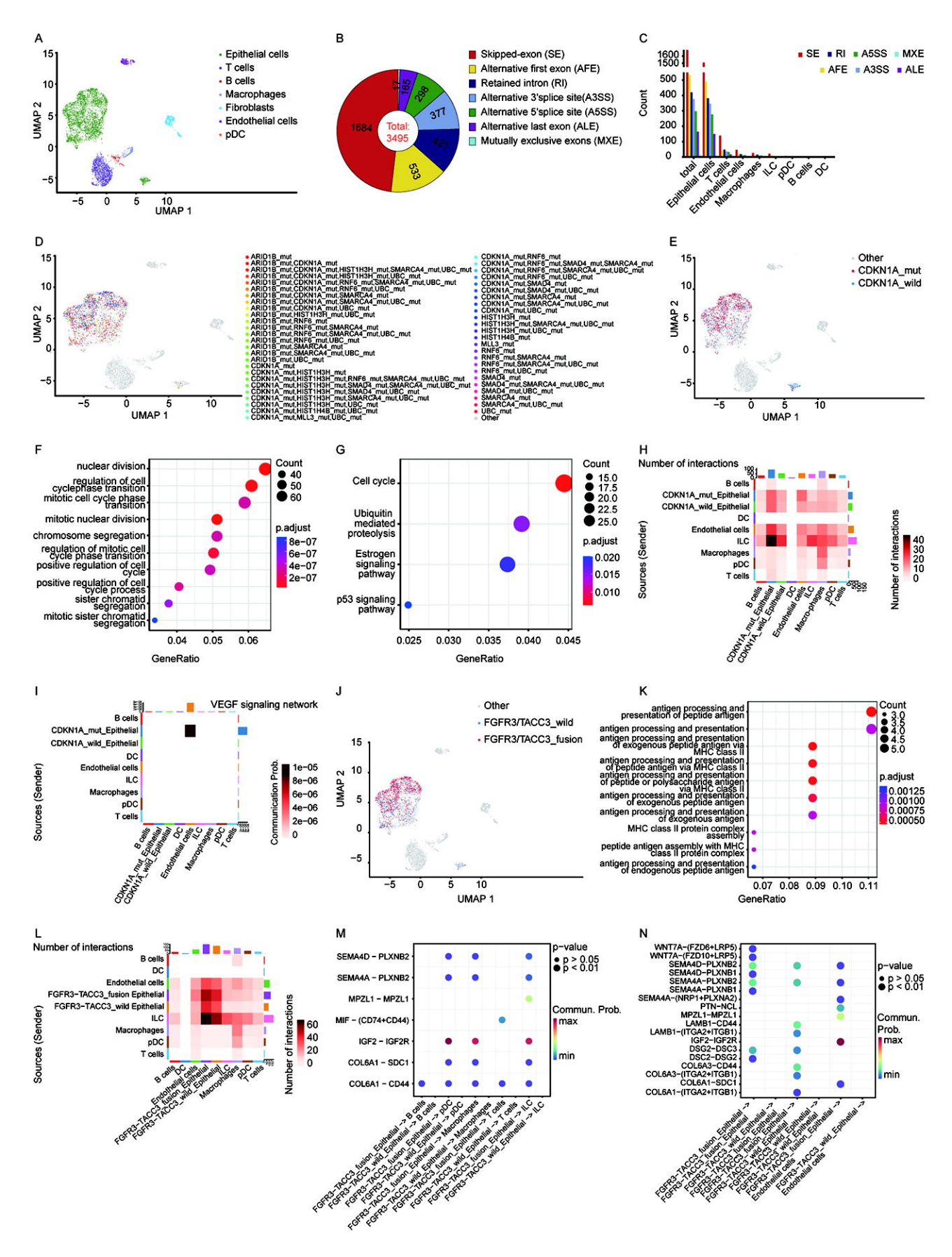
[21]. Jiang, H., et al., The role of semaphorin 4D in tumor development and angiogenesis in human breast

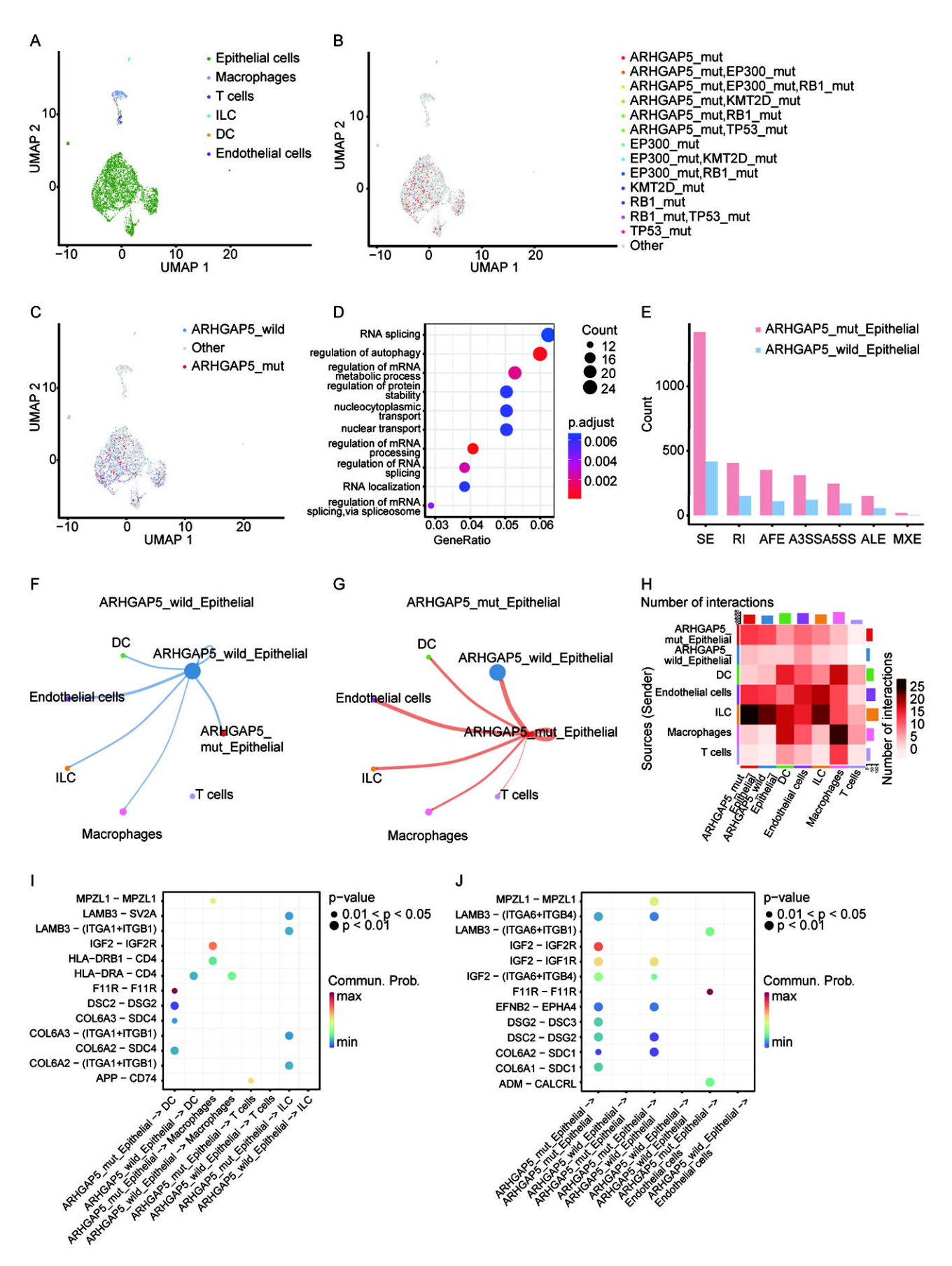
cancer. Onco Targets Ther, 2016. 9: p. 5737-5750.
[22]. Zhang, X., et al., High expression of COL6A1 predicts poor prognosis and response to immunotherapy in bladder cancer. Cell Cycle, 2023. 22(5): p. 610-618.
[23]. Wang, A., et al., MPZL1 Promotes Lung Adenocarcinoma Progression by Enhancing Tumor Proliferation, Invasion, Migration, and Suppressing Immune Function via Transforming Growth Factor-beta1. Hum Gene Ther, 2023. 34(11-12): p. 540-553.
[24]. Chen, D., L. Cao and X. Wang, MPZL1 promotes tumor cell proliferation and migration via activation of Src kinase in ovarian cancer. Oncol Rep, 2019. 42(2): p. 679-687.
[25]. Feng, J., et al., MPZL1 upregulation promotes tumor metastasis and correlates with unfavorable prognosis in non-small cell lung cancer. Carcinogenesis, 2022. 43(10): p. 919-929.
[26]. Huang, X., et al., Wnt7a activates canonical Wnt signaling, promotes bladder cancer cell invasion, and is suppressed by miR-370-3p. J Biol Chem, 2018. 293(18): p. 6693-6706.
[27]. Cera, M.R., et al., Increased DC trafficking to lymph nodes and contact hypersensitivity in junctional adhesion molecule-A-deficient mice. J Clin Invest, 2004. 114(5): p. 729-38.
[28]. Leech, A.O., et al., Paradigms lost-an emerging role for over-expression of tight junction adhesion proteins in cancer pathogenesis. Ann Transl Med, 2015. 3(13): p. 184.
[29]. Snyder, A., et al., Contribution of systemic and somatic factors to clinical response and resistance to PD-L1 blockade in urothelial cancer: An exploratory multi-omic analysis. PLoS Med, 2017. 14(5): p. e1002309.
[30]. Braun, D.A., et al., Interplay of somatic alterations and immune infiltration modulates response to PD-1 blockade in advanced clear cell renal cell carcinoma. Nat Med, 2020. 26(6): p. 909-918.
[31]. 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.
[32]. Tang, F., et al., mRNA-Seq whole-transcriptome analysis of a single cell. Nat Methods, 2009. 6(5): p. 377-82.
[33]. Chen, H., et al., Pan‐Cancer Single‐Nucleus Total RNA Sequencing Using snHH‐Seq. Advanced Science, 2023.
[34]. Marzluff, W.F., E.J. Wagner and R.J. Duronio, Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail. Nat Rev Genet, 2008. 9(11): p. 843-54.


Cancer specimens
A
3'scRNA-seq
Cell Loading
Bead Loading
Cell Lysis & Barcoding
Reverse Transcription
B
Library amplification
intermediate cDNA library
fluorescence
DNA length
Next-generation sequencing
cell identity
UMAP2
UMAP1
Chigene V1
C
mRNA full-length labeling
Library amplification
Sequencing Library
wild type
Fusion
gene A
gene B
Mutation
Indel/SNV
Splicing
exon 7
exon 9
cell identity
UMAP2
UMAP1
Modified Barcode oligo
microbead
cell barcoding oligo strcture
DNA extension
template switch oligo
annealing site
TTTTT ..... TTTTTT
AAAAA ..... AAAAAA
cDNA sythesis with template switch oligo
cDNA
mRNA
Random bridging ligation & extension
microbead
exon 1
exon 2
exon 3
exon 4
exon 5
exon 6
exon 7
exon 9
Splicing
exon n
mutation
Bridge oligo 1
Bridge oligo 2
Gene A cDNA first strand
exon on gene A
gene Splicing site
Gene B cDNA first strand
mutution spot on gene A
Molecular feature enrichment
P5
cell barcode
UMI
5'
mutation/fusion
cDNA
3'
P7
Integrated analysis
UMAP2
UMAP1
Chigene V2


A
Capture efficiency
Doublet rate
Percentage(%)
0
20
40
60
5000
10000
20000
30000
40000
50000
Loaded cell number
B
Cell
C
Bead
Cell
D
Bead Occupancy Rate (%)
0
25
50
75
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
Specific 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
B cells
CD4 T cells
CD8 T cells
Monocytes
Neutrophils
NK cells
Mast cells
platelets
Percent
Chigene V1
10x
BD
singleron
Sample
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
B cells
DC
Endothelial cells
Epithelial cells
Fibroblasts
Macrophages
Mast cells
Plasma cells
T cells
Percent
Chigene V1
10X
Sample


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
Specific 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
B cells
DC
Endothelial cells
Epithelial cells
Fibroblasts
Macrophages
Mast cells
Plasma cells
T cells
pDC
Percent
Chigene V1
Chigene V2
Sample


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
Sources (Sender)
B cells
CDKN1A_mut_Epithelial
CDKN1A_wild_Epithelial
DC
Endothelial cells
ILC
Macrophages
pDC
T cells
Macro-phages
I
VEGF signaling network
Communication Prob.
J
Other
FGFR3/TACC3_wild
FGFR3/TACC3_fusion
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)
FGFR3−TACC3_fusion Epithelial
FGFR3−TACC3_wild Epithelial
M
SEMA4D − PLXNB2
SEMA4A − PLXNB2
MPZL1 − MPZL1
MIF − (CD74+CD44)
IGF2 − IGF2R
COL6A1 − SDC1
COL6A1 − CD44
p−value
p > 0.05
p < 0.01
Commun. Prob.
max
min
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)


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
Endothelial cells
ILC
Macrophages
T cells
G
ARHGAP5_mut_Epithelial
DC
Endothelial cells
ILC
Macrophages
T cells
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