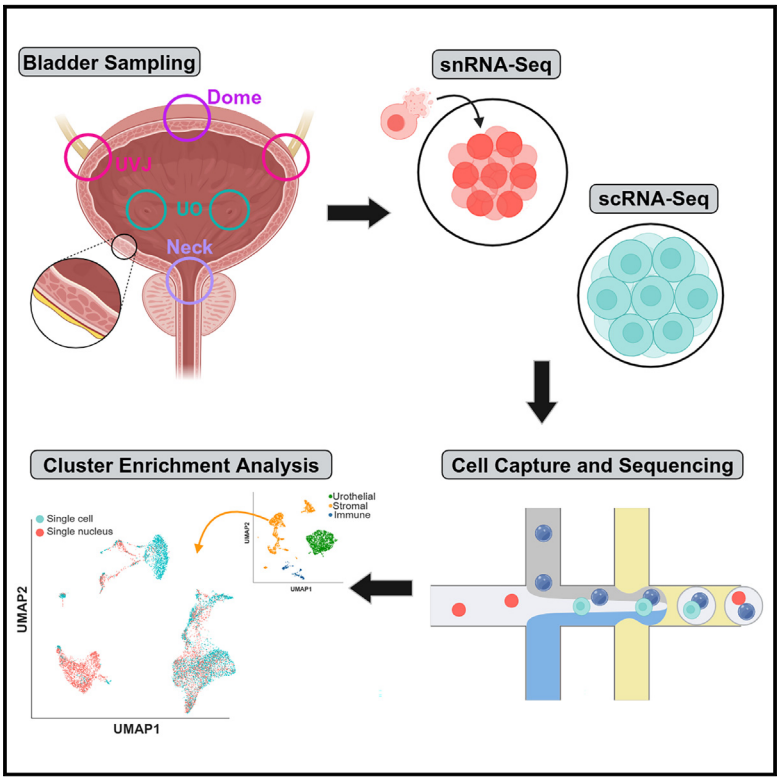
iScience

Article

Exploring the utility of snRNA-seq in profiling human bladder tissue: A comprehensive comparison with scRNA-seq

Graphical abstract



Highlights

- Multiple bladder anatomical regions were sampled to capture cell type diversity
- Joint analysis of scRNA-seq and snRNA-seq was completed to compare technologies
- Non-coding RNA enrichment in snRNA-seq data obscured cell type identification
- SnRNA-seq approach could capture rare cell types and better resolve cell clusters

Authors

Briana Santo, Emily E. Fink, Alexandra E. Krylova, ..., Oliver Wessely, Byron H. Lee, Angela H. Ting

Correspondence

ahting@mdanderson.org

In brief

Biological sciences; Cell biology; Transcriptomics

Santo et al., 2025, iScience *28*, 111628
January 17, 2025
https://doi.org/10.1016/j.isci.2024.111628

CellPress

iScience

CellPress
OPEN ACCESS

Article

Exploring the utility of snRNA-seq in profiling human bladder tissue: A comprehensive comparison with scRNA-seq

Briana Santo,[1] Emily E. Fink,[2,3] Alexandra E. Krylova,[1] Yi-Chia Lin,[4] Mohamed Eltemamy,[4] Alvin Wee,[4] Oliver Wessely,[5] Byron H. Lee,[6] and Angela H. Ting[1,7,*]

[1]Epigenetics & Molecular Carcinogenesis, M.D. Anderson Cancer Center, Houston, TX 77054, USA
[2]Genomic Medicine, Lerner Research Institute, Cleveland Clinic, Cleveland, OH 44195, USA
[3]Charles River Laboratories, Garfield Heights, OH 44128, USA
[4]Department of Urology, Glickman Urological and Kidney Institute, Cleveland Clinic, Cleveland, OH 44195, USA
[5]Department of Cardiovascular & Metabolic Sciences, Lerner Research Institute, Cleveland Clinic, Cleveland, OH 44195, USA
[6]Department of Urology, M.D. Anderson Cancer Center, Houston, TX 77054, USA
[7]Lead contact
*Correspondence: ahting@mdanderson.org
https://doi.org/10.1016/j.isci.2024.111628

SUMMARY

Single cell sequencing technologies have revolutionized our understanding of biology by mapping cell diversity and gene expression in healthy and diseased tissues. While single-cell RNA sequencing (scRNA-seq) has been widely used, interest in single-nucleus RNA sequencing (snRNA-seq) is growing due to its benefits, including the ability to analyze archival tissues and capture rare cell types that are challenging to dissociate. However, comparative studies across tissues have yielded mixed results, with some reporting enhanced cell type retention using snRNA-seq while others finding cell type identification to be challenging in snRNA-seq data. The GUDMAP consortium aims to construct a molecular atlas of the lower urinary tract (LUT); thus, we set out to determine the strengths and limitations of each approach in characterizing LUT cell types. Using the human bladder, we determined that scRNA-seq offered more discriminative gene sets for identification while snRNA-seq could facilitate capture of previously underrepresented cell types.

INTRODUCTION

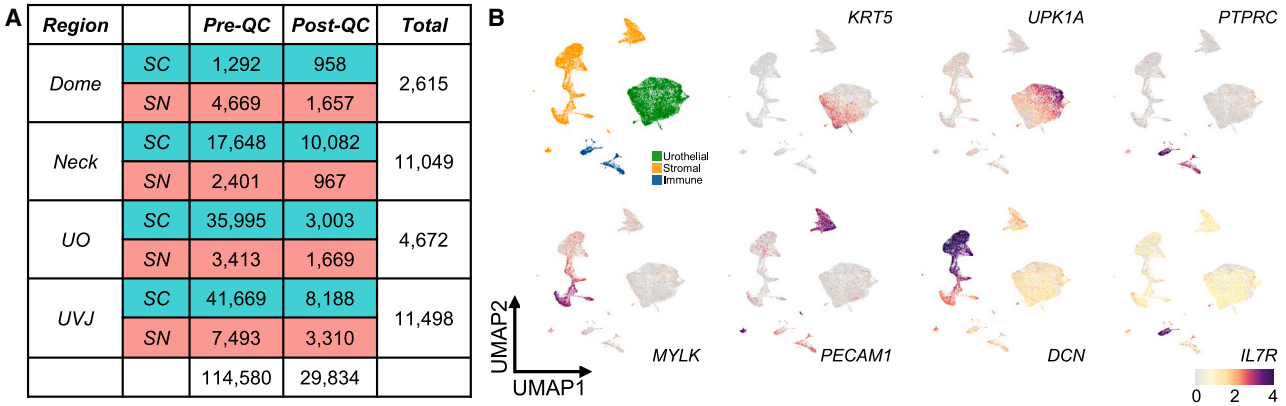
Single cell sequencing technologies have revolutionized our ability to interrogate biology in health and disease, enabling the mapping of cell type diversity and expression signatures within and across organ systems.[1] While most studies to date have used scRNA-seq for molecular mapping of tissues, an increasing number of studies are exploring snRNA-seq profiling. Their reason for doing so is motivated by the proposed benefits of snRNA-seq, which include the ability to analyze archival tissues, thus increasing statistical power, and the potential to capture rare cell types in robust numbers.[2,3]

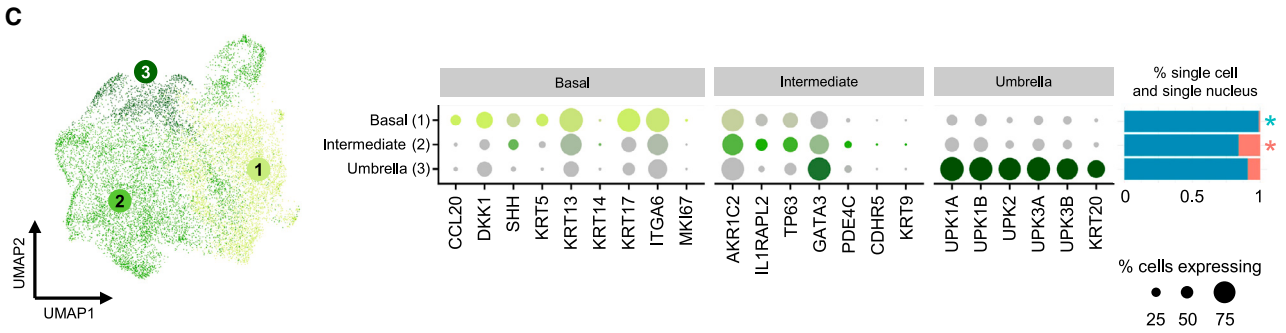
Select studies have demonstrated that snRNA-seq facilitates the capture of cells, such as glia and specialized epithelia, that often evade robust quantification by snRNA-seq.[1,2,4–7] The low yield for such cell types has been attributed to their delicate cell structure, which may not survive single cell dissociation protocol. Meanwhile, the isolation of nuclei for snRNA-seq is more feasible for such cell types. To determine whether snRNA-seq profiling will yield more robust quantification of a tissue's cellular composition, some studies have completed head-to-head comparisons of scRNA-seq and snRNA-seq cell type capture. Currently, comparative studies have profiled human kidney, brain, lung, breast, and PBMCs.[2,4–8] Findings are mixed, with kidney studies reporting significant increase in podocyte retention,[4,8] while studies on breast report challenges in cell type identification.[7] Through such studies, a concern has arisen—the enrichment of non-coding RNAs in differentially expressed genes[7] and the reduction in canonical markers. As a consequence, cell type identification in single-nucleus exclusive studies, particularly in disease or when studying new biology, may prove to be a challenge. Thus, to establish which data type is best suited for the molecular profiling of human anatomy, it is essential to complete comparative studies for most if not all organ and tissue types.

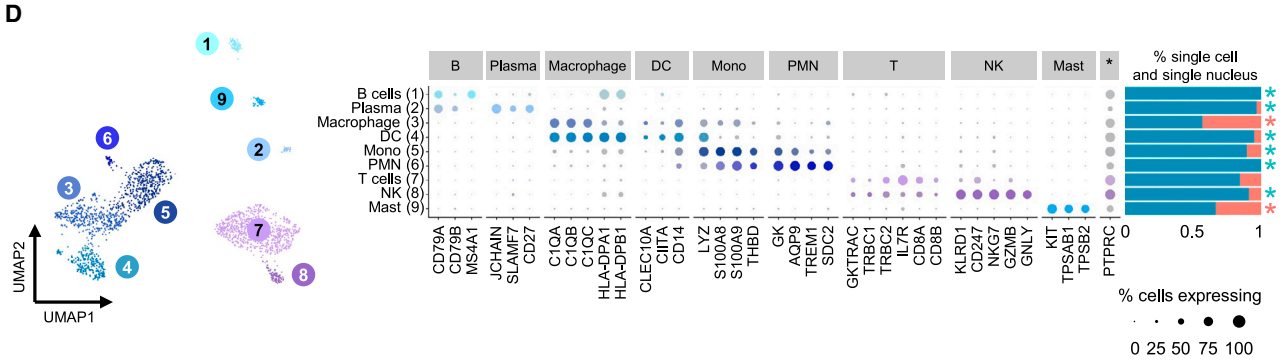
Robust data collection is essential to ensure that generated reference datasets are sensible, meaningful, and practical resources for the research community. Toward the objectives of the Genitourinary Development Molecular Anatomy Project (GUDMAP) consortium, which in its current funding cycle aims to construct a molecular anatomy reference of the lower urinary tract (LUT) in human and model organism,[9,10] we set out to compare scRNA-seq and snRNA-seq for their ability to capture and characterize LUT cell types. We chose to focus on the human bladder given its complexity in cellular composition and regional anatomical diversity.[11] Our hypothesis was that while,

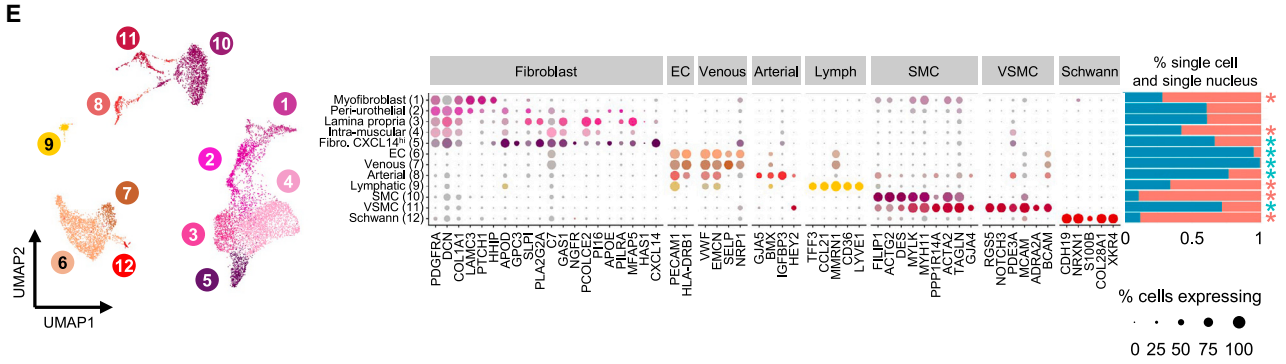
A

Region		Pre-QC	Post-QC	Total
Dome	SC	1,292	958	2,615
	SN	4,669	1,657	
Neck	SC	17,648	10,082	11,049
	SN	2,401	967	
UO	SC	35,995	3,003	4,672
	SN	3,413	1,669	
UVJ	SC	41,669	8,188	11,498
	SN	7,493	3,310	
		114,580	29,834	









(legend on next page)

for specific cell types, snRNA-seq would provide for better enrichment, scRNA-seq would provide more discriminative gene sets for cell type identification and functional annotation. To test our hypothesis, we collected, sequenced, and computationally integrated single-cell and single-nucleus samples from paired specimens originating from each of the four major bladder regions [dome, neck, ureteral orifice (UO), and ureterovesical junction (UVJ)][12] and compared each cell type's single-cell and single-nucleus subpopulations based on differentially expressed genes and enriched biological processes.

RESULTS

Single-cell and single-nucleus profiling of human bladder cells

To interrogate single-cell and single-nucleus technologies for their ability to profile bladder cell types, we performed joint single-cell and single-nucleus analysis of cells collected from the dome, neck, UO, and UVJ of human bladder tissue. Histologically normal human bladder tissue was procured from a Donation after Brain Death organ donor. The use of tissues from a single donor helped to minimize gene expression variability and allow for a true head-to-head comparison between scRNA-seq and snRNA-seq. Using multiple different bladder regions further added to the robustness of this comparison. Following tissue collection, samples were immediately dissociated into single-cell suspensions and processed for scRNA-seq. In parallel, an adjacent specimen from each bladder region was frozen and batch-processed for snRNA-seq later. The average sequencing depth across samples was 540.8M (Table S1). After quality control and filtering, a total of 29,834 cells comprised of 22,231 single cells and 7,603 single nuclei (Figure 1A), from four human bladder regions each (dome, neck, UO, UVJ), were segregated into 14 clusters by unsupervised clustering (Figure S1A; Table S2). Samples from each bladder region and data type (single-cell and single-nucleus) were uniformly distributed throughout the UMAP (Figure S1B). To partition UMAP clusters into the three major cellular compartments of human bladder—urothelium, stroma, and immune—we visualized established markers for each compartment as feature plots (Figure 1B). For the urothelium, we previewed *KRT5*, a marker for uroepithelial basal cells, and *UPK1A*, a marker for umbrella cells.[13,14] We found that clusters 0, 1, 4, and 12 comprised a urothelial island where the gradient from basal to umbrella expression could be discerned based on plotted makers. For the immune compartment, we plotted *PTPRC*, the leukocyte common antigen, and *IL7R*, a marker of lymphoid and developing immune cells,[14,15] thus identifying clusters 7 and 8 as the immune compartment. Lastly, to identify stromal clusters, we plotted *MYLK* for smooth muscle cells, *PECAM1* for endothelial cells, and *DCN* for fibroblasts.[15] As a result, we could assign all remaining clusters (2,3,5,6,9–11,13) to the stromal compartment. A comprehensive dot plot visualizing canonical markers across pertinent bladder cell types was also constructed and evaluated toward compartment assignment (Figure S1C).

Classification of the urothelial cells

Re-analysis of the urothelial subset resulted in three clusters (Figure 1C, UMAP). The differential gene expression analysis (DEA) showed that each cluster corresponded to one of the general urothelial layers: the basal urothelium (*KRT5*$^+$/*KRT13*$^+$/*KRT17*$^+$/*MKI67*$^+$), the umbrella cells (*UPK1A*$^+$/*UPK1B*$^+$/*UPK2*$^+$/*UPK3A*$^+$/*UPK3B*$^+$/*KRT20*$^+$), and the intermediate or transitional urothelium (*AKR1C2*$^+$/*IL1RAPL2*$^+$/*TP63*$^+$/*GATA3*$^+$/*CDHR5*$^+$) (Figure 1C, Dot Plot).[13,16] The intermediate urothelial cluster was the largest of the three clusters (8,108 cells), followed by the basal (5,179 cells) and the umbrella (1,489 cells) cell populations. Interestingly, the intermediate cell population showed significant single nuclei enrichment (16.1%), while the umbrella (90.7%) and the basal (98.9%) populations were made up almost entirely of single cells (Figure 1C, Bar Plot).

Classification of the immune cells

Re-clustering of the immune subset resulted in nine clusters (Figure 1D, UMAP). Following differential gene expression analysis, we identified one cluster each of B Cells (*CD79A*$^+$/*CD79B*$^+$/*MS4A1*$^+$), plasma cells (*JCHAIN*$^+$/*SLAMF7*$^+$/*CD27*$^+$), macrophage (*C1QA*$^+$/*C1QB*$^+$/*C1QC*$^+$), dendritic cells (DC) (*CLEC10A*$^+$/*CIITA*$^+$), monocytes (*S100A8*$^+$/*S100A9*$^+$), neutrophils (*GK*$^+$/*AQP9*$^+$/*TREM1*$^+$), T cells (*Il7R*$^+$/*CD8A*$^+$/*CD8B*$^+$), natural killer (NK) cells (*KLRD1*$^+$/*CD247*$^+$/*GZMB*$^+$), and mast cells (*KIT*$^+$/*TPSAB1*$^+$/*TPSB2*$^+$) (Figure 1D, Dot Plot).[15,17] The T cell cluster was the largest of the immune compartment with 736 cells, while the smallest was the plasma cell cluster with 31 cells. Interestingly, macrophage (42.5%) and mast (32.7%) cells demonstrated significant single-nucleus enrichment, while plasma (96.8%), DC (95.0%), monocyte (89.6%), neutrophil (100%), and NK cells (91.1%) had significant single-cell enrichment. T cells, the largest immune cell population, were not enriched significantly for either data type (Figure 1D, Bar Plot).

Classification of the stromal cells

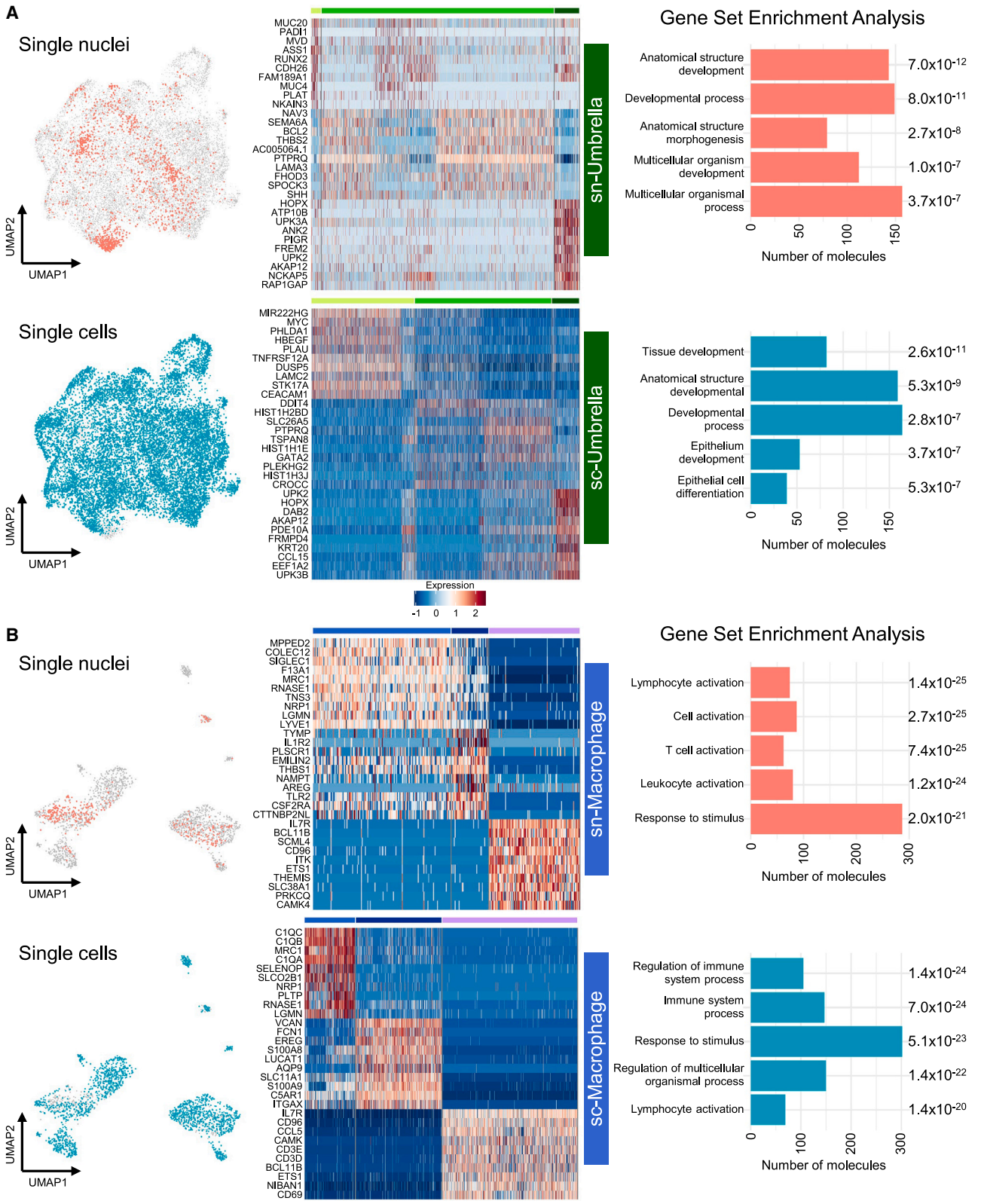
Analysis of the stromal subset resulted in twelve clusters (Figure 1E, UMAP). For each stromal cluster, established markers were distinctly expressed within DEG sets, facilitating unambiguous assignment of cell types. As expected, several distinct fibroblast clusters (*PDGFRA*$^+$/*DCN*$^+$/*COL1A1*$^+$) were observed. In this case, we identified five clusters which we refer to as myofibroblast (*MYLK*$^+$/*MYH11*$^+$), peri-urothelial

Figure 1. Joint single-cell and single-nucleus profiling of human bladder cells

(A) Distribution of single cells and single nuclei per bladder region before and after quality control filtering.

(B) UMAP with assigned bladder compartments and feature plots for established urothelial, stromal, and immune markers that facilitated compartment assignment.

(C–E) UMAPs of the urothelial (C), immune (D), and stromal (E) compartments based on gene expression. Adjacent to the UMAPs, dot plots for each corresponding compartment show the expression level of known markers across cell clusters identified in each compartment. Far right bar plots show the enrichment of single cells and single nuclei in each compartment. *Asterisks indicate statistically significant enrichment.



(legend on next page)

fibroblast (*LAMC3*$^{+}$/*PTCH1*$^{+}$/*APOE*$^{+}$), lamina propria fibroblast (*GAS1*$^{+}$/*PCOLCE2*$^{+}$/*PI16*$^{+}$/*MFAP5*$^{+}$), intra-muscular fibroblast (*C7*$^{+}$/*PCOLCE2*$^{-}$), and CXCL14^{hi} fibroblast (*GPC3*$^{+}$/*NFGR*$^{+}$/*CXCL14*$^{+}$).[14,15,18,19] In addition, several distinct endothelial and smooth muscle cell clusters were identified. The endothelial cell population comprised a more general endothelial cluster (*PECAM1*$^{+}$/*GJA5*$^{-}$/*SELP*$^{-}$/*LYVE1*$^{-}$), that was not enriched for either arterial or venous markers exclusively, a venous cluster (*SELP*$^{+}$/*NRP1*$^{+}$), an arterial cluster (*GJA5*$^{+}$/*BMX*$^{+}$), and a lymphatic cluster (*LYVE1*$^{+}$/*TFF3*$^{+}$/*CD36*$^{+}$).[14] With respect to smooth muscle cells, we identified two distinct clusters corresponding to general smooth muscle cells (*ACTA2*$^{+}$/*ACTG2*$^{+}$/*DES*$^{+}$) and vascular smooth muscle cells (*RGS5*$^{+}$/*NOTCH3*$^{+}$/*MCAM*$^{+}$).[14,19] Lastly, we identified a single Schwann cell cluster (*CDH19*$^{+}$/*NRXN*$^{+}$/*XKR4*$^{+}$).[20]

Like the urothelial and immune compartments, we assessed the enrichment of single cells and single nuclei for each stromal cell type. Of the twelve clusters, the largest was the general Endothelial cell cluster, comprising 2,485 cells, and 94.6% of which were procured by the single-cell protocol. Other cell types demonstrating single-cell enrichment were CXCL14^{hi} fibroblast (66.4%), venous endothelial (99.2%), arterial endothelial (76.5%), and vascular smooth muscle cell clusters (71.6%). The next largest population was smooth muscle cells, where 88.6% of cells were of single-nucleus origin. In addition to the smooth muscle cell cluster, significant single-nucleus enrichment was observed in the myofibroblast (71.5%), intra-muscular fibroblast (57.7%), lymphatic endothelial (65.8%), and Schwann (87.5%) cell clusters. Peri-urothelial and lamina propria fibroblast clusters did not show significant enrichment of a particular data type.

Assessment of single-cell and single-nucleus expression profiles in the urothelium

To evaluate whether single-nucleus data would produce discriminative cell markers for cell type identification, we subset each urothelial cluster into its single-cell and single-nucleus subpopulations and completed DEA comparing each cluster subpopulation (e.g., basal single-nucleus) against all other urothelial cluster subpopulations (e.g., intermediate and basal single-nucleus subpopulations) (Tables S3 and S4). Then, we plotted the top ten differentially expressed genes (DEGs) for each cluster subpopulation on a heatmap (Figure 2A, Heatmaps). For the single-nucleus cluster heatmap, only the top markers for the umbrella population aligned with established urothelial markers (e.g., *UPK2* and *UPK3A*) and the top DEGs did not distinguish the basal, intermediate, and umbrella cells well. This observation was particularly apparent for the intermediate cells, where top markers included non-coding RNAs (e.g., *AC005064.1*) and the heatmap showed highly variable expression among the intermediate cells. In stark comparison, the top markers based on single-cell data included canonical urothelial markers (e.g., *UPK2*, *UPK3B*, *KRT20*, and *CEACAM1*) and demonstrated cell-type specific patterns of DEG expression for the basal, intermediate, and umbrella cell populations. To further evaluate how well DEGs computed from single-nucleus and single-cell subpopulations could help decipher cell types and functions, we performed gene set enrichment analysis using each cell subpopulation's DEG set. It was found that DEGs computed from single-nucleus subpopulations corresponded to very general biological terms (Table S6), whereas those computed from single-cell subpopulations correctly indicated epithelial cells (Table S7). For example, umbrella single-nucleus DEGs highlighted "Anatomical structure development," "Developmental process," and "Multicellular organism development" as enriched terms; although in practice these are very generic biological processes that could apply to any cell type. Meanwhile, umbrella single-cell DEGs suggested pertinent processes such as "Epithelium development" and "Epithelial cell differentiation" (Figure 2A, Gene Ontology Bar Plot).

Assessment of single-cell and single-nucleus expression profiles in the immune cells

As done for the urothelial compartment, we evaluated how well single-nucleus and single-cell top DEGs differentiated each immune cell type, and which biological processes were enriched for all DEGs. Once again, single-cell top marker expression provided for better separation of immune clusters. These patterns were particularly apparent for the macrophage, monocyte, and T cell populations (Figure 2B, Heatmaps). For single-nucleus top DEGs, notable markers included *LYVE1* and *RNASE1* for macrophage, *TLR2* for monocytes, and *IL7R* for T cells. Meanwhile, canonical markers were prominent in single-cell top DEGs for macrophage (*C1QA*, *C1QB*, *C1QC*), monocytes (*S100A8*, *S100A9*), and T cells (*IL7R*, *CD3D*, *CD3E*, *CD69*). Enriched biological processes were comparable between datatypes, with single-nucleus DEGs highlighting "Lymphocyte activation," "T cell activation," and "Leukocyte activation,"

Figure 2. Assessment of single-cell and single-nucleus gene expression in the bladder urothelial and immune cell populations

(A) UMAPs show the urothelial compartment partitioned by single-nucleus and single-cell instances. Heatmaps show the top ten differentially expressed genes (DEGs) for each urothelial cell type. The upper heatmap shows DEGs for urothelial cell type's single-nucleus subpopulations, while the lower heatmap shows those DEGs computed for single-cell subpopulations. Differential expression analysis was completed by comparing a urothelial cell type's single-nucleus or single-cell subpopulation against all other cell type's respective subpopulations. The color bar above each heatmap indicates which heatmap entries belong to each urothelial cell type (basal, intermediate, or umbrella, color coded the same as in Figure 1C UMAP). Bar plots show enriched biological processes for the umbrella cell population based on all single-nucleus or single-cell differentially expressed genes.

(B) UMAPs show the immune compartment based on gene expression and partitioned by single-nucleus and single-cell instances. Heatmaps show the top ten differentially expressed genes (DEGs) for each immune cell type with sufficient nuclei and cells for differential expression analysis (macrophage, monocyte, and T cells). The upper heatmap shows DEGs for immune cell type's single-nucleus subpopulations while the lower heatmap shows those DEGs computed for single-cell subpopulations. Differential expression analysis was completed by comparing an immune cell type's single-nucleus or single-cell subpopulation against all other cell type's respective subpopulations. The color bar above each heatmap indicates which heatmap entries belong to each immune cell type (macrophage, monocyte, and T cells, color coded the same as in Figure 1D UMAP). Bar plots show enriched biological processes for the macrophage cell population based on all single-nucleus or single-cell differentially expressed genes.

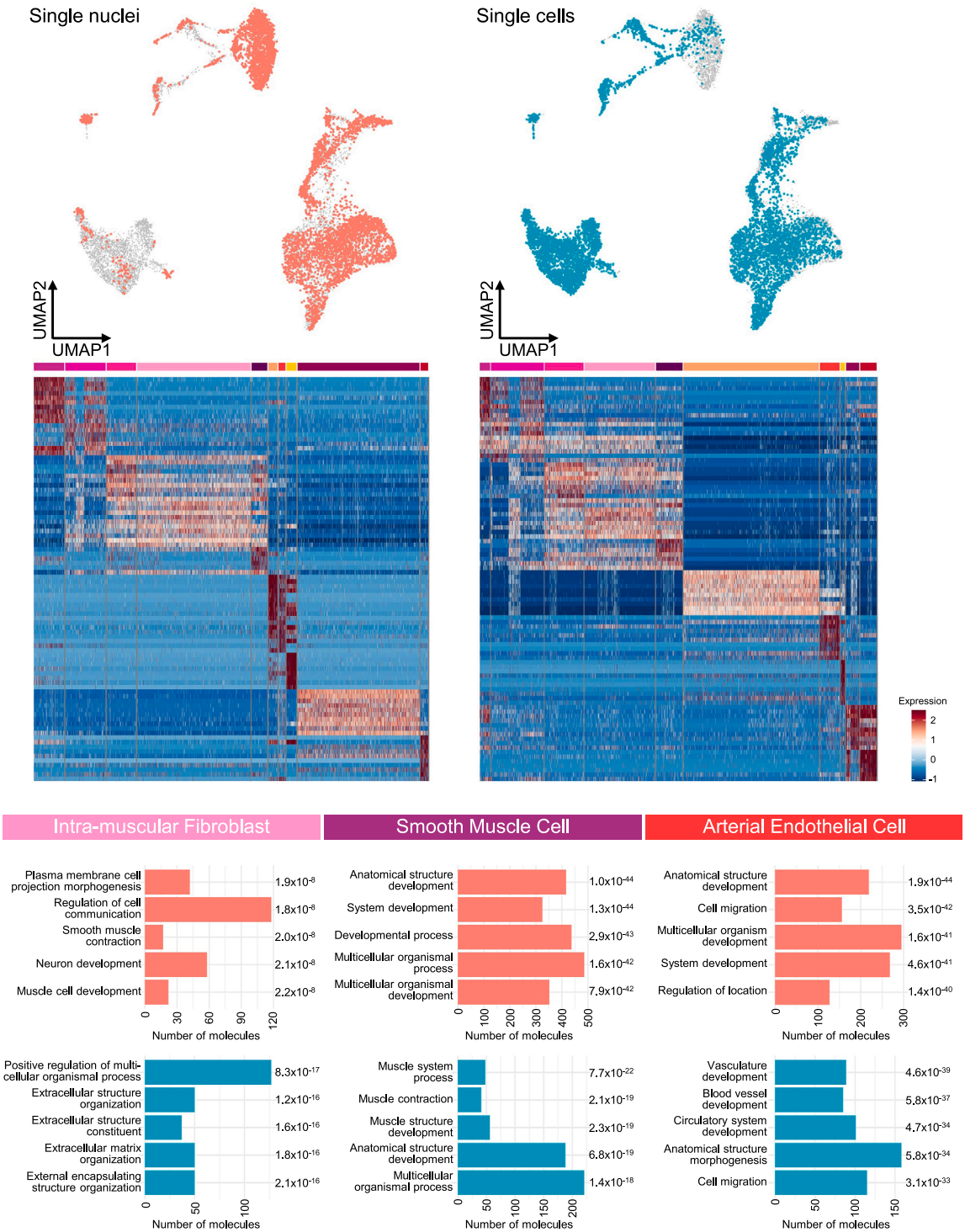


Figure 3. Assessment of single-cell and single-nucleus enrichment in the stromal bladder cell populations

UMAPs show the stromal compartment based on gene expression and partitioned by single-nucleus and single-cell instances. Heatmaps show the top ten differentially expressed genes (DEGs) for each stromal cell type with sufficient nuclei and cells for differential expression analysis (myofibroblast, peri-urothelial fibroblast, lamina propria fibroblast, intra-muscular fibroblast, CXCL14^{hi} fibroblast, smooth muscle cells, vascular smooth muscle cells, as well as general, arterial, and venous endothelial cells). The left heatmap shows DEGs for stromal cell type's single-nucleus subpopulations while the right heatmap shows those

(legend continued on next page)

and single-cell DEGs highlighting ''Regulation of immune system process,'' ''Immune system process,'' and ''Lymphocyte activation'' (Figure 2B, Gene Ontology Bar Plot). Overall, in the immune compartment, we observed less of a discrepancy between data types in their ability to help decipher cell types.

Assessment of single-cell and single-nucleus expression profiles in the stromal cells

Finally, when we evaluated how well single-nucleus and single-cell top DEGs differentiated stromal cell types, separation among cell types appeared comparable between single-nucleus and single-cell top DEGs (Figure 3, Heatmaps; Figures S1D and S1E). However, upon further inspection, the single-nucleus data showed enrichment for non-coding RNAs in the top 10 DEGs for several cell types while canonical cell markers were absent (Table S3). For example, the smooth muscle cell single-nucleus subpopulation top DEGs included non-coding RNAs *CHRM3* and *AC079313.2*, but lacked canonical markers such as *ACTA2*, *ACTG2*, or *MYH11*. These canonical markers were, however, found in the smooth muscle cell single-cell subpopulation, facilitating unambiguous cell type identification. Similar observations were made for the fibroblast and vascular smooth muscle cell populations, where canonical markers such as *PDGFRA* and *RGS*, respectively, were absent in top single-nucleus DEGs but present in single-cell top DEGs (Table S4). A formal Chi-squared analysis found that non-coding RNAs were significantly enriched in the single nucleus DEG list of arterial and lymphatic endothelial cells, all fibroblast subpopulations, smooth muscle cells, and vascular smooth muscle cells (Table S5).

The difference between single-nucleus and single-cell DEGs was even more apparent in the gene ontology analysis. Like our analysis of the urothelial compartment, single-nucleus DEG sets indicated very generic biological processes (Figure 3, Gene Ontology Bar Plots). For example, for the intra-muscular fibroblast, smooth muscle cell, and arterial endothelial cell populations, top terms included ''Neuron development,'' ''Developmental process,'' and ''Multicellular organism development,'' respectively. In stark contrast, single-cell DEG sets produced pertinent terms including ''Extracellular structure organization'' for intra-muscular fibroblast, ''Muscle contraction'' for smooth muscle cells, and ''Blood vessel system development'' for arterial endothelial cells. Subsequently ranked gene ontology terms demonstrated a similar trend, with vague, general biology terms coming from single-nucleus DEG sets (Table S6) and cell-type specific and correctly associated terms coming from single-cell DEG sets (Table S7).

Since we had robust numbers of both cells and nuclei in the stromal compartment, we additionally performed independent, parallel analyses of each data source. In the single-nucleus analysis, we identified 13 stromal clusters (Figures S2A and S2B), including an additional fibroblast subtype (Fibroblast-6) and an additional VSMC resembling pericytes (VSMC2; positive for *CSPG4*). Conversely, in the single-cell analysis, we identified 9 cell clusters (Figures S3A and S3B) and observed attrition of the intramuscular fibroblast subpopulation and loss of resolution between SMC and VSMC cells. These results echoed the findings from the single cell/single nucleus proportions test in the joint analysis (Figure 1E). When visualizing the top cell cluster markers, we also noted a higher number of non-coding RNAs in the single-nucleus analysis (Figure S2C) when compared with the single-cell analysis (Figure S3C).

DISCUSSION

Here, we describe the joint analysis of scRNA-seq and snRNA-seq data from human donor bladder tissues. We procured tissues from the dome, neck, UO, and UVJ anatomical regions and generated scRNA-seq and snRNA-seq data from paired samples of each region from a single donor. To properly integrate these two data types, we completed extensive quality control filtering and batch correction, leveraging an established Seurat workflow.[21] Upon cell type identification, we further evaluated which cell types demonstrated single-nucleus enrichment over single-cell. For each cell population comprising a mix of single cells and single nuclei, we completed gene set enrichment analysis separately for the single-cell and single-nucleus subpopulations to see what biological processes were highlighted by each cell type's single-cell and single-nucleus gene sets. In doing so, we demonstrated, through a data-driven approach, an anticipated challenge with snRNA-seq data; the difficulty to decipher cell type based on top marker genes and enriched biological processes alone. While joint analysis of scRNA-seq and snRNA-seq facilitated cell identification in our study, studies using single-nucleus data alone, and especially those studies featuring discovery of *de novo* cell types, face the added challenge of deciphering non-coding RNA-predominated cluster markers and vague nucleus-oriented ontology terms.

To date, few studies have completed a head-to-head comparison of scRNA-seq and snRNA-seq data, and none have done so using paired samples from human bladder tissues. Prior studies have shown that the use of snRNA-seq would facilitate the capture and interrogation of cell types that have historically been difficult to capture using single-cell protocols. For example, cells such as glia, kidney podocytes, and lung alveolar epithelia and fibroblasts are difficult to isolate in robust numbers during tissue dissociation for scRNA-seq.[2,4–6] The inability to isolate such cell types has negative implications for studies where those cell types are the major drivers of disease pathology, such as alveolar epithelia and fibroblasts in lung fibrosis.[2] Ongoing studies, including ours, aim to identify which cell types in which organs and anatomical regions would be best quantified by snRNA sequencing. Bormann et al.[22] and Wang et al.[23] demonstrated how single-nucleus technology enables collection of glia in high volume toward identification of cell and molecular signatures associated with oligodendrocyte and astrocyte response

DEGs computed for single-cell subpopulations. Differential expression analysis was completed by comparing a stromal cell type's single-nucleus or single-cell subpopulation against all other cell type's respective subpopulations. The color bar above each heatmap indicates which heatmap entries belong to each stromal cell type (color coded the same as in Figure 1E). Bar plots show enriched biological processes for the intra-muscular fibroblast, smooth muscle cell, and arterial endothelial cell populations based on all single-nucleus or single-cell differentially expressed genes.

to ischemic injury in stroke and glioblastoma recurrence.[5,6] The authors suggested that the morphology of glia hinders their isolation by single-cell methods. Similarly, podocytes, which are known for their distinct and fragile morphology, are often underrepresented in single-cell studies. In a recent work, Wu et al.[24] reported a 20-fold increase in podocyte retention for samples processed by snRNA-seq.[8] Like lung epithelia, fibroblasts, and glia, podocyte quantification is essential to evaluate renal disease pathology.[4,8]

The advantages of snRNA-seq are not limited to enrichment for rare cell types. Single-nucleus technology can be readily applied to cryopreserved samples, enabling researchers to tap into expansive and archival biospecimen collections and thus has become the mainstay for some of the large profiling consortiums such as the Kidney Precision Medicine Project (KPMP).[1–3] Additionally, snRNA-seq has been found to significantly reduce dissociation-induced transcriptional stress responses in cells, which have the potential to confound findings in pathological states or cells with distinct mitochondrial signatures.[3,8] snRNA-seq has also helped reduce noise in the transcriptional profiling of macrophage, which due to their phagocytic function may contain other cell type's transcriptomes when quantified by scRNA-seq. Despite these benefits, the enrichment of non-coding RNAs cannot be ignored. When working with large scale snRNA-seq datasets, it is important to have context for cell types in disease, informed by prior scRNA-seq studies, so that non-coding RNA enrichment does not obscure cell type identification. As noted by Kumar et al., such non-coding RNA enrichment confounded canonical marker enrichment for common epithelial cell types.[7] Approaches to manage non-coding RNA enrichment include filtering out non-coding RNAs prior to sample integration and batch correction so that highly variable genes used in clustering are protein-coding exclusively, or as done in our study, integrate and co-analyze scRNA-seq and snRNA-seq datasets.

In our study, the joint analysis of snRNA-seq and scRNA-seq datasets allowed us to identify cell types with ease as well as demonstrate the benefits and disadvantages of single-nucleus data for human bladder tissue. Across the urothelial, immune, and stromal compartments, we identified several cell types which showed significant single-nucleus enrichment. In the urothelial compartment, we found that intermediate urothelial cells demonstrated significant single nucleus enrichment. Despite this enrichment, intermediate urothelial as well as basal and umbrella cell single-nucleus subpopulations were difficult to identify based on top markers and gene ontology. Meanwhile, urothelial cell types identified by scRNA-seq showed significant canonical marker and epithelial process enrichment. In the immune compartment, we found macrophage and mast cells to have significant single-nucleus enrichment. Interestingly, the top differentially expressed genes and gene ontology terms derived from single-nucleus and single-cell immune subpopulations were comparable. Identification of immune cell type was possible using either the single-nucleus or the single-cell DEGs and biological processes.

The greatest discrepancy between single-nucleus and single-cell data was seen in the stromal compartment. While several cell types demonstrated significant single-nucleus enrichment, including some fibroblasts, smooth muscle cells, and lymphatic endothelial cells, top marker gene expression and ontology analysis were inadequate for cell type identification. For example, canonical markers for these cell types were significantly reduced in single-nucleus DEG sets. In addition, enriched biological processes were either too vague for cell type identification (e.g., anatomical structure and system development for smooth muscle cells) or indicated the incorrect cell type (e.g., smooth muscle contraction for intra-muscular fibroblasts). When these same cell types were investigated by single-cell analysis, canonical markers were well enriched in the top ten DEGs per cluster, ontology terms were pertinent (e.g., muscle contraction and muscle structure development for smooth muscle cells), and overall separation among cell types based on gene expression data was clear in heatmaps. In the stromal compartment, the greatest advantage conferred by single-nucleus data was the collection of a sizable Schwann cell population. Prior studies report little enrichment of neuronal cell types when profiling the LUT, likely due to the morphology of these cells.[14]

When we further compared the single-nucleus and single-cell data in independent analyses, additional benefits of snRNA-seq became apparent. When single-nucleus data was analyzed by itself, increased resolution of cell type subpopulations were observed, including additional fibroblast and vascular smooth muscle cell populations. We also observed a non-coding RNA enrichment in the top DEGs per stromal cell type. These observations are in stark contrast to the independent single-cell analysis, wherein we observed decreased cell type resolution. Limited sample size notwithstanding, it cannot be ignored that using non-coding RNA enriched single nucleus data may facilitate biological insights from the human bladder that would otherwise be overlooked by purely single cell data. Interactive visualizations of all processed data with cell and data type annotations are available at the NIH-CZI CELLxGENE Discover Cellular Visualization Tool.

Limitations of the study

In this work, we only had eight samples from a single human donor. To ensure comprehensive profiling of cell type diversity in both healthy reference datasets and disease states, it is essential that additional studies analyze paired scRNA-seq and snRNA-seq data with a larger sample size. Our samples were, however, taken from each of the four major bladder regions, with paired sampling for single-nucleus and single-cell data generation, providing a balanced and anatomically diverse sampling of the organ. Although we lacked spatial transcriptomic validation of our findings, all canonical markers used for cell type identification in this study are referenced from prior works in the kidney, bladder, ureter, and other regions of the LUT.[12–14,18] We also recognize that current gene ontology databases are largely based on coding genes and thus biased against non-coding RNAs. However, single-nucleus studies such as ours present unique opportunities to augment existing knowledgebases with non-coding RNA and biological process associations. In summary, our results here helped to establish a framework for sample preparation and data collection and analysis for human bladder tissues in both sexes and across the lifespan to generate a LUT cell atlas as part of GUDMAP.

RESOURCE AVAILABILITY

Lead contact

Correspondence and requests for materials should be addressed to the lead contact, Angela H. Ting (ahting@mdanderson.org).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- Raw and processed sequencing data have been deposited in SRA and the Gene Expression Omnibus and are publicly available as of the date of publication. Accession numbers are listed in the key resources table. In addition, interactive visualizations of all processed data with cell type and data type annotations are available at the NIH-CZI CELLxGENE Discover - Cellular Visualization Tool and can be accessed here.
- All original code has been deposited on GitHub and is publicly available.
- Any additional information required to reanalyze the data reported in this paper can be requested from the lead contact.

ACKNOWLEDGMENTS

We acknowledge the support of Lifebanc in procuring bladder tissues from the deceased organ donor for this research. This work is supported by U01 DK131383 to O.W., B.H.L., and A.H.T. National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) funds U01 DK131383.

AUTHOR CONTRIBUTIONS

Conceptualization, B.H.L. and A.H.T.; formal analysis, B.S. and A.E.K.; resources, B.H.L., Y.L., M.E., and A.W.; investigation, B.S., E.E.F., and A.E.K.; software, B.S.; supervision, A.H.T.; writing – original draft, B.S., B.H.L., and A.H.T.; writing – review and editing, E.E.F., Y.L., M.E., A.W., and O.W.; funding acquisition, O.W., B.H.L., and A.H.T.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENT MODEL AND STUDY PARTICIPANT DETAILS
  - Human bladder procurement
- METHOD DETAILS
  - Definition of bladder regions
  - Single cell isolation
  - Single nucleus isolation
  - Sample processing and library preparation
  - scRNA-seq and snRNA-seq data joint analysis
  - scRNA-seq and snRNA-seq subset analysis
  - scRNA-seq v. snRNA-seq gene set analysis
  - Independent scRNA-seq and snRNA-seq analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2024.111628.

Received: June 3, 2024
Revised: November 1, 2024
Accepted: December 16, 2024
Published: December 18, 2024

REFERENCES

1. Ding, J., Adiconis, X., Simmons, S.K., Kowalczyk, M.S., Hession, C.C., Marjanovic, N.D., Hughes, T.K., Wadsworth, M.H., Burks, T., Nguyen, L.T., et al. (2020). Systematic comparison of single-cell and single-nucleus RNA-sequencing methods. Nat. Biotechnol. *38*, 737–746.
2. Koenitzer, J.R., Wu, H., Atkinson, J.J., Brody, S.L., and Humphreys, B.D. (2020). Single-nucleus RNA-sequencing profiling of mouse lung. Reduced dissociation bias and improved rare cell-type detection compared with single-cell RNA sequencing. Am. J. Respir. Cell Mol. Biol. *63*, 739–747.
3. Kim, N., Kang, H., Jo, A., Yoo, S.-A., and Lee, H.-O. (2023). Perspectives on single-nucleus RNA sequencing in different cell types and tissues. J. Pathol. Transl. Med. *57*, 52–59.
4. Lake, B.B., Chen, S., Hoshi, M., Plongthongkum, N., Salamon, D., Knoten, A., Vijayan, A., Venkatesh, R., Kim, E.H., Gao, D., et al. (2019). A single-nucleus RNA-sequencing pipeline to decipher the molecular anatomy and pathophysiology of human kidneys. Nat. Commun. *10*, 2832.
5. Bormann, D., Knoflach, M., Poreba, E., et al. (2024). Single-nucleus RNA sequencing reveals glial cell type-specific responses to ischemic stroke in male rodents. Nat Commun *15*, 6232. https://doi.org/10.1038/s41467-024-5.
6. Wang, L., Jung, J., Babikir, H., Shamardani, K., Jain, S., Feng, X., Gupta, N., Rosi, S., Chang, S., Raleigh, D., et al. (2022). A single-cell atlas of glioblastoma evolution under therapy reveals cell-intrinsic and cell-extrinsic therapeutic targets. Nat. Can. (Ott.) *3*, 1534–1552.
7. Kumar, T., Nee, K., Wei, R., He, S., Nguyen, Q.H., Bai, S., Blake, K., Pein, M., Gong, Y., Sei, E., et al. (2023). A spatially resolved single-cell genomic atlas of the adult human breast. Nature *620*, 181–191.
8. Wu, H., Kirita, Y., Donnelly, E.L., and Humphreys, B.D. (2019). Advantages of single-nucleus over single-cell RNA sequencing of adult kidney: rare cell types and novel cell states revealed in fibrosis. J. Am. Soc. Nephrol. *30*, 23–32.
9. Harding, S.D., Armit, C., Armstrong, J., Brennan, J., Cheng, Y., Haggarty, B., Houghton, D., Lloyd-MacGilp, S., Pi, X., Roochun, Y., et al. (2011). The GUDMAP database–an online resource for genitourinary research. Development *138*, 2845–2853.
10. McMahon, A.P., Aronow, B.J., Davidson, D.R., Davies, J.A., Gaido, K.W., Grimmond, S., Lessard, J.L., Little, M.H., Potter, S.S., Wilder, E.L., et al. (2008). GUDMAP: the genitourinary developmental molecular anatomy project. J. Am. Soc. Nephrol. *19*, 667–671.
11. Abedini, A., Zhu, Y.O., Chatterjee, S., Halasz, G., Devalaraja-Narashimha, K., Shrestha, R., S Balzer, M., Park, J., Zhou, T., Ma, Z., et al. (2021). Urinary single-cell profiling captures the cellular diversity of the kidney. J. Am. Soc. Nephrol. *32*, 614–627.
12. Liaw, A., Cunha, G.R., Shen, J., Cao, M., Liu, G., Sinclair, A., and Baskin, L. (2018). Development of the human bladder and ureterovesical junction. Differentiation *103*, 66–73.
13. Yu, Z., Liao, J., Chen, Y., Zou, C., Zhang, H., Cheng, J., Liu, D., Li, T., Zhang, Q., Li, J., et al. (2019). Single-cell transcriptomic map of the human and mouse bladders. J. Am. Soc. Nephrol. *30*, 2159–2176.
14. Fink, E.E., Sona, S., Tran, U., Desprez, P.-E., Bradley, M., Qiu, H., Eltemamy, M., Wee, A., Wolkov, M., Nicolas, M., et al. (2022). Single-cell and spatial mapping Identify cell types and signaling Networks in the human ureter. Dev. Cell *57*, 1899–1916.e6.
15. Chen, Z., Zhou, L., Liu, L., Hou, Y., Xiong, M., Yang, Y., Hu, J., and Chen, K. (2020). Single-cell RNA sequencing highlights the role of inflammatory cancer-associated fibroblasts in bladder urothelial carcinoma. Nat. Commun. *11*, 5077.
16. McConkey, D.J., Lee, S., Choi, W., Tran, M., Majewski, T., Lee, S., Siefker-Radtke, A., Dinney, C., and Czerniak, B. (2010). Molecular genetics of bladder cancer: Emerging mechanisms of tumor initiation and progression. Urol. Oncol. *28*, 429–440.

17. Zhong, J., Ding, R., Jiang, H., Li, L., Wan, J., Feng, X., Chen, M., Peng, L., Li, X., Lin, J., et al. (2022). Single-cell RNA sequencing reveals the molecular features of peripheral blood immune cells in children, adults and centenarians. Front. Immunol. *13*, 1081889.

18. Apodaca, G. (2023). Defining the molecular fingerprint of bladder and kidney fibroblasts. Am. J. Physiol. Ren. Physiol. *325*, F826–F856.

19. Joseph, D.B., Henry, G.H., Malewska, A., Reese, J.C., Mauck, R.J., Gahan, J.C., Hutchinson, R.C., Malladi, V.S., Roehrborn, C.G., Vezina, C.M., and Strand, D.W. (2021). Single-cell analysis of mouse and human prostate reveals novel fibroblasts with specialized distribution and microenvironment interactions. J. Pathol. *255*, 141–154.

20. Wei, Z., Shu, S., Zhang, M., Xie, S., Tang, S., Nie, K., and Li, H. (2021). A subpopulation of Schwann cell-like cells with nerve regeneration signatures is identified through single-cell RNA sequencing. Front. Physiol. *12*, 637924.

21. Sona, S., Bradley, M., and Ting, A.H. (2023). Protocols for single-cell RNA-seq and spatial gene expression integration and interactive visualization. STAR Protoc. *4*, 102047.

22. Bormann, D., Knoflach, M., Poreba, E., Riedl, C.J., Testa, G., Orset, C., Levilly, A., Cottereau, A., Jauk, P., Hametner, S., et al. (2024). Single-nucleus RNA sequencing reveals glial cell type-specific responses to ischemic stroke in male rodents. Nat. Commun. *15*, 6232. https://doi.org/10.1038/s41467-024-50465-z.

23. Wang, L., Jung, J., Babikir, H., Shamardani, K., Jain, S., Feng, X., Gupta, N., Rosi, S., Chang, S., Raleigh, D., et al. (2022). A single-cell atlas of glioblastoma evolution under therapy reveals cell-intrinsic and cell-extrinsic therapeutic targets. Nat. Cancer *3*, 1534–1552. https://doi.org/10.1038/s43018-022-00475-x.

24. Wu, H., Kirita, Y., Donnelly, E.L., and Humphreys, B.D. (2019). Advantages of single-nucleus over single-cell RNA sequencing of adult kidney: rare cell types and novel cell states revealed in fibrosis. J. Am. Soc. Nephrol. *30*, 23–32. https://doi.org/10.1681/ASN.2018090912.

25. Satija, R., Farrell, J.A., Gennert, D., Schier, A.F., and Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. Nat. Biotechnol. *33*, 495–502.

26. Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M., and Trombetta, J.J. (2015). Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. Cell *161*, 1202–1214.

27. Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., 3rd, Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive Integration of Single-Cell Data. Cell *177*, 1888–1902. https://doi.org/10.1016/j.cell.2019.05.031.

28. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., and Hoffman, P. (2021). Integrated analysis of multimodal single-cell data. Cell *184*, 3573–3587.

29. Miller, S.A., Policastro, R.A., Sriramkumar, S., Lai, T., Huntington, T.D., Ladaika, C.A., Kim, D., Hao, C., Zentner, G.E., and O'Hagan, H.M. (2021). LSD1 and aberrant DNA methylation mediate persistence of enteroendocrine progenitors that support BRAF-mutant colorectal cancer. Cancer Res. *81*, 3791–3805.

30. Zappia, L., and Oshlack, A. (2018). Clustering trees: a visualization for evaluating clusterings at multiple resolutions. GigaScience *7*, giy083.

31. Patterson-Cross, R.B., Levine, A.J., and Menon, V. (2021). Selecting single cell clustering parameter values using subsampling-based robustness metrics. BMC Bioinf. *22*, 39.

32. Kolberg, L., Raudvere, U., Kuzmin, I., Vilo, J., and Peterson, H. (2020). gprofiler2–an R package for gene list functional enrichment analysis and namespace conversion toolset g: Profiler. F1000Research *9*, ELIXIR-709.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Human bladder tissues from brain dead donor patients collected through LifeBanc.	This paper	NA
Chemicals, peptides, and recombinant proteins		
ACK Lysis Buffer	Thermo Fisher Scientific	Cat#A1049201
SPRIselect	Beckman	Cat#B23317
Tissue-Tek® O.C.T. Compound	Sakura	Cat#4583
Critical commercial assays		
Papain Dissociation System	Worthington	Cat#LK003150
Chromium Next GEM Single Cell 3' Kit v3.1	10X Genomics	PN-1000268
Chromium Next GEM Chip G Single Cell Kit	10X Genomics	PN-1000120
Chromium i7 Multiplex Kit	10X Genomics	PN-120262
Library Construction Kit	10X Genomics	PN-1000196
Dual Index Plate TT Set A	10X Genomics	PN-3000431
Agilent High Sensitivity DNA Kit Reagents	Agilent	Cat#5067-4626
QuantiTect SYBR Green PCR Kit	Qiagen	Cat#204145
Deposited data		
Single-Cell and Single-Nucleus RNA-seq FASTQs	This paper	PRJNA1111560
Processed single-cell and single-nucleus RNA-seq data	This paper	GEO: GSE267964
Processed single-cell and single-nucleus RNA-seq data And interactive visualization	This paper	https://cellxgene.cziscience.com/collections/0e54d4de-44f0-4d50-8649-b5c2bbe8f5d1
Software and algorithms		
Codes forked for Seurat data integration and clustering pipeline	GitHub; Fink et al.[14]	https://github.com/basanto/scRNA_Analysis_Developing_for_Human_Bladder_Analysis
Codes used for differential expression and gene ontology analysis	This paper; GitHub	https://github.com/basanto/Human_Bladder_scRNA-seq_vs_snRNA-seq; Zenodo [https://doi.org/10.5281/zenodo.14446321]
R	CRAN	https://cran.r-project.org/
Seurat (4.3.0.1)	Satija et al.[25]; Macosko et al.[26]; Stuart et al.[27]; Hao et al.[28]	https://cran.r-project.org/web/packages/Seurat/index.html
cellranger (7.1.0)	10X Genomics	https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/7.1/release-notes
clustree (0.5.1)	CRAN	https://cran.r-project.org/web/packages/clustree/index.html
argparser (0.7.1)	CRAN	https://cran.r-project.org/web/packages/argparse/index.html
data.table (1.15.2)	CRAN	https://cran.r-project.org/web/packages/data.table/index.html
dplyr (1.1.4)	CRAN	https://cran.r-project.org/web/packages/dplyr/index.html
factoextra (1.0.7)	CRAN	https://cran.r-project.org/web/packages/factoextra/index.html

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
ggplot2 (3.5.0)	CRAN	https://cran.r-project.org/web/packages/ggplot2/index.html
magrittr (2.0.3)	CRAN	https://cran.r-project.org/web/packages/magrittr/index.html
parallelly (1.37.1)	CRAN	https://cran.r-project.org/web/packages/parallelly/index.html
stringr (1.5.0)	CRAN	https://cran.r-project.org/web/packages/stringr/index.html
tibble (3.2.1)	CRAN	https://cran.r-project.org/web/packages/tibble/index.html
tidyr (1.3.0)	CRAN	https://cran.r-project.org/web/packages/tidyr/index.html
scProportionTest (0.0.0.9000)	GitHub; Miller et al.[29]	https://github.com/rpolicastro/scProportionTest
scCustomize (2.1.2)	CRAN	https://cran.r-project.org/web/packages/scCustomize/index.html
gprofiler2 (0.2.3)	CRAN	https://cran.r-project.org/web/packages/gprofiler2/index.html
ggcorrplot (0.1.4.1)	CRAN	https://cran.r-project.org/web/packages/ggcorrplot/readme/README.html
htmlwidgets (1.6.4)	CRAN	https://cran.r-project.org/web/packages/htmlwidgets/index.html
reshape2 (1.4.4)	CRAN	https://cran.r-project.org/web/packages/reshape2/index.html
patchwork (1.2.0)	CRAN	https://cran.r-project.org/web/packages/patchwork/index.html
qs (0.26.1)	CRAN	https://cran.r-project.org/web/packages/qs/index.html
janitor (2.2.0)	CRAN	https://cran.r-project.org/web/packages/janitor/index.html
future (1.33.0)	CRAN	https://cran.r-project.org/web/packages/future/index.html
ggraph (2.1.0)	CRAN	https://cran.r-project.org/web/packages/ggraph/index.html
lattice (0.21.8)	CRAN	https://cran.r-project.org/web/packages/lattice/index.html
gridExtra (2.3)	CRAN	https://cran.r-project.org/web/packages/gridExtra/index.html
grid (4.1.0)	CRAN	https://cran.r-project.org/web/packages/grid/index.html
Other		
Flowmi™ Cell Strainers	Bel-Art	Cat#136800040

EXPERIMENT MODEL AND STUDY PARTICIPANT DETAILS

Human bladder procurement

Donor tissue was made available through a research collaboration with Lifebanc, a non-profit organization that coordinates organ recovery for use in transplantation in more than 80 hospitals in Northeast Ohio. Under our agreement, Lifebanc screens potential donors for suitability, including checking the applicable donor registry to determine if a potential donor has made a donor designation as recognized under applicable state law. With family consent, Lifebanc offers the LUT including the bladder to us for research procurement. In this study, one bladder was procured from a 21-year-old male donor, and eight different samples were excised, two from each major bladder region (Dome, Neck, Ureteral Orifice, Ureterovesical Junction). One sample from each bladder region was used for single cell dissociation, and the other sample was used for single nucleus isolation.

METHOD DETAILS

Definition of bladder regions

To comprehensively sample bladder cell types and diversity in embryologic origin, we collected tissue samples from the bladder dome, neck, ureteral orifice (UO), and ureterovesical junction (UVJ). For the purposes of this study, and based on the literature, the bladder regions were defined as follows. The bladder dome, or apex, was obtained from the anterosuperior region of the bladder where the medial umbilical ligament enters. The dome represents the distensible part of the bladder, which accommodates urine during bladder filling and expels urine during micturition. The bladder neck was obtained from the confluence of the bladder and urethra inferior to the trigone. It acts to funnel urine into the urethra during voiding and contains the internal urethral sphincter, which involuntarily contracts during storage and relaxes during emptying. The ureteral orifice was obtained from the visible opening of the ureter into the bladder. The ureterovesical junction was taken from the area superolateral to the ureteral orifice, which contains the intramural segment of ureter and a stroma-rich anti-reflux mechanism.

Single cell isolation

Under sterile conditions (working entirely within BSL2 Biosafety cabinets), freshly dissected human LUT tissues from each designated anatomic location are immediately washed with PBS and centrifuged for 5 min at 1000xg. The PBS is aspirated off, and the tissues are transferred to a sterile Petri dish where they are subsequently minced using sterile razor blades until a "paste-like" consistency is achieved (<5 min). The tissue digestion is carried out using the Papain Dissociation System (Worthington Biochemical Corporation) according to the manufactures instructions with minor modifications that were empirically determined and optimal for human ureter and bladder tissues. Briefly, 5 mL of sterile Earle's Balanced Salt Solution (EBSS) is added to one vial of papain and equilibrated at 37°C, and 500 μL EBSS is added to one vial of DNase and placed on ice. The minced tissue is transferred to a 50 mL conical containing 5 mL papain and 500 μL DNase for incubation with shaking (70 rpm) at 37°C for 1h. Using a P1000 pipette, the tissue is gently pipetted up and down to check for proper digestion. 5.5 mL of ovoid inhibitor is then added to the tissue suspension to stop the digestion. The mixture is filtered by passing the entire volume (~11 mL) through a 70 μm cap filter to remove large debris. The cell suspension is subsequently centrifuged at 1000rpm for 5 min to pellet cells, and the supernatant is discarded. The cell pellet is gently re-suspended in 3 mL ACK lysis buffer and incubated at room temperature for 3 min to remove red blood cells. The cells are then centrifuged at 1000rpm for 5 min, washed with another 5 mL PBS, and re-suspended in a volume of PBS appropriate for the cell pellet size. The re-suspended cells are then passed through a 40 μm FLOWMI cell strainer and counted with trypan blue to note cell viability. The final cell concentration is recorded. An aliquot of cell suspension for 10X GEM generation is then prepared according to the cell suspension volume calculator table of the Chromium NextGEM Single Cell 3′ Reagent Kits v3.1 user guide.

Single nucleus isolation

Dissected bladder regions were embedded in optimal cutting temperature (O.C.T.) compound, snap frozen with dry ice, and stored at −80°C until use. We followed the protocol established by the Kidney Precision Medicine Consortium (KPMP) for nuclei isolation.[4] Briefly, we cut ten 40 μm-thick cryosections and placed tissue sections into 1 mL of ice-cold nuclear extraction buffer (NEB, 20 mM Tris pH 8.0, 329 mM sucrose, 5 mM $CaCl_2$, 3 mM $MgAc_2$, 0.1 mM EDTA, 0.1% Triton X-100 with 0.1% RNase Inhibitor). Using a p1000 pipette with the tip trimmed to increase bore size, the tissue mixture was pipetted up and down at least 20 times to dissolve the O.C.T. The mixture was further mixed by pipetting up and down 10 times using a regular, untrimmed p1000 pipette and then transferred to a Dounce homogenizer on ice. The tissue was homogenized on ice using 5 strokes with pestle A followed by 20 strokes with pestle B, taking care to minimize bubble formation. The homogenized mixture was incubated on ice for 10 min and then passed through a 40 μm FLOWMI cell strainer into a new 15 mL conical tube. The sample volume was brought up to 10 mL with ice-cold PBSE (1x PBS, 1mL EGTA) and centrifuged at 900 g for 10 min at 4°C to pellet the nuclei. Finally, the nuclei were resuspended in ice-cold PBSE containing 1% BSA for manual counting.

Sample processing and library preparation

For both single cell and single nucleus data, we used 10X Genomics Chromium Single Cell 3′ Reagents Kit version 3.1. We targeted to recover 10,000 cells/nuclei per sample unless a given sample had fewer than 10,000 total cells/nuclei. All samples had >5,000 cells/nuclei captured for library generation and sequencing. Following the 10X protocol, we added the single-cell/single-nucleus suspension, the gel beads, and the emulsion oil to the 10X Genomics Single Cell Chip G and ran the Chromium Controller. Immediately following the droplet generation, samples were transferred to a PCR 8-tube strip (USA scientific), and reverse transcription was performed using SimpliAmp thermal cycler (Applied Biosystems). Following reverse transcription, cDNA was recovered using the recovery reagent provided by 10X Genomics. The cDNA was cleaned up using the Silane DynaBeads according to the 10X Genomics user guide. The purified cDNA was amplified for 11 cycles and subsequently cleaned up using SPRIselect beads (Beckman Coulter). To determine the cDNA concentrations, 1:10 dilution of each sample was analyzed on an Agilent Bioanalyzer High Sensitivity chip. The cDNA libraries were constructed according to the Chromium Single Cell 3′ Reagent Kit version 3.1 user guide.

scRNA-seq and snRNA-seq data joint analysis

For both scRNAseq and snRNAseq samples, libraries were pooled and sequenced to a target depth of 50,000 read pairs per cell. De-multiplexed FASTQ files were processed with CellRanger (v7.1.0), where reads were mapped using the count pipeline with the pre-built reference genome refdatagex-GRCh38-2020A and GTF from GENECODE v32 (GRCh28.p13). Summary statistics for each sample's alignment performance have been provided in Table S1. The downstream analysis was performed using Seurat 4.3.0.1 in R[25] along with an established R analysis pipeline.[21] The raw data, comprising of 114,580 cells (n = 96,604 cells and n = 17,976 nuclei), was filtered for low-quality cells using QC thresholds determined by assessing the distribution plots. Briefly, cells with mitochondrial reads >25% of total mapped reads, gene counts <500 or >12,000, and total mapped read counts <4,000 were filtered out, leaving 40,097 high-quality cells (n = 31,764 cells and n = 8,333 nuclei) for the downstream analysis. Seurat's standard workflow was followed. Data were integrated, or batch corrected, using CCA and the IntegrateData functionality. Next, using the 2,000 most variable genes, principal components (PCs) were computed, and the first 30 PCs were utilized to generate clustering at a resolution 0.25. Cell clusters were visualized using UMAP and annotated by marker gene expression. Cluster level similarity was assessed by calculating cluster correlations (Pearson's) based on the top 2000 variable genes, which grouped the clusters into 3 major compartments – urothelial, stromal, and immune. Subset analysis for each compartment was performed using the same Seurat workflow. Clustering optimization was performed using clustree and chooseR.[30,31]

scRNA-seq and snRNA-seq subset analysis

For each subset analysis, cells from clusters assigned to a given compartment (immune, stromal, or urothelial) were processed by the same analysis pipeline as described above, starting from raw gene counts. The subsets were subjected to additional filtering and re-processing until any cross-compartment contamination was sufficiently removed. Potentially contaminating clusters (e.g., suspected doublets) were removed 'If the top100 differential marker list of a given cluster shows markers that should be exclusively expressed by cells of another compartment (e.g., PDGFRA appears among top100 markers in an immune subset cluster)' AND 'the contaminating marker is expressed in >10% of cells in that cluster'. If such parameters exist, the subset was clustered using these parameters and subjected to the same subset analysis workflow, while excluding the contaminating cluster(s). This resulted in 2,094 immune cells (n = 1,725 cells and n = 369 single nuclei), 12,964 stromal cells (n = 7,227 cells and n = 5,737 single nuclei), and 14,776 urothelial cells (n = 13,279 cells and n = 1,497 single nuclei). Each subset data was then clustered using 50 PCs and a resolution of 0.4 for immune, 40 PCs and a resolution of 0.3 for stromal, and 25 PCs and a resolution of 0.1 for urothelial.

For each compartment (immune, stromal, and urothelial), cell types were called using canonical and established markers. The proportion of single cell and single nucleus enrichment for each cell cluster was then computed and statistical significance ($p < 0.05$) was assessed using scProportionTest.[29] A summary of cell counts, single-cell and single-nucleus enrichment, and differential expression analysis for each cell type is provided in Table S2.

scRNA-seq v. snRNA-seq gene set analysis

When both single cells and single nuclei were present in sufficient numbers ($n > 50$ each), we aimed to assess the differential gene expression of these two data types for each captured cell type. For each cell type cluster, when applicable, the single-cell and single-nucleus subpopulations were subset, and the differential gene list for each was found using Seurat's FindMarkers function. Here, the single-cell (or single-nucleus) subpopulation's expression was compared against the expression of the single-cell (or single-nucleus) subpopulation of all other cell types within the compartment (e.g., immune). The upregulated markers were filtered in each list using a threshold of Log2FC > 1.5 and adjusted p-value (q) < 0.05. Gene Ontology (GO) enrichment analysis for each cluster's single-cell and single-nucleus gene list was performed using gprofiler2, and gene ontology biological process, molecular functions, and cell components were recorded.[32] We annotated each gene in Tables S3 and S4 with its GENCODE biotype information (https://www.gencodegenes.org/pages/biotypes.html). To determine whether non-coding RNA enrichment was statistically significant in a particular cell and data type, we completed chi-square tests for association between two categorical variables, where the 2x2 table was comprised of the number of non-coding RNAs and protein coding genes for a specific cell type compared between single cell and single nucleus gene lists (Table S5).

Independent scRNA-seq and snRNA-seq analysis

We analyzed the single cell and single nucleus data independently following identical analysis workflow as described above. Briefly, the aligned single-cell or single-nucleus data were filtered for high quality cells/nuclei, integrated, and clustered with optimization. Clusters belonging to the stromal, immune, and urothelial compartments were identified, and the cells/nuclei belonging to each compartment were subset. For each compartment, cells/nuclei were reintegrated, and clustering optimization was completed. Only the stromal compartment contained robust numbers of both cells and nuclei for meaningful cell type identification and differential gene expression analysis.

QUANTIFICATION AND STATISTICAL ANALYSIS

The proportionality of single cell or single nucleus subpopulations was assessed using scProportionTest, and significance was defined at FDR <0.05 and abs(Log2FC) > 0.58. Differential expression analysis was completed using the Seurat FindMarker function set to the default Wilcoxon Rank-Sum test. Significant differentially expressed genes were identified by filtering for an abs(Log2FC) > 1.5 and q < 0.05. Chi-square tests for association between two categorical variables were used to determine whether non-coding RNA enrichment was statistically significant in a particular cell and data type, with significance set at $p < 0.05$.