Nat Biotechnol. Author manuscript; available in PMC 2021 October 10.

Published in final edited form as:
Nat Biotechnol. 2021 October 01; 39(10): 1239–1245. doi:10.1038/s41587-021-00926-3.

Revealing the spatio-phenotypic patterning of cells in healthy and tumor tissues with mLSR-3D and STAPL-3D

Ravian L. van Ineveld[1,2,#], **Michiel Kleinnijenhuis**[1,2,#], **Maria Alieva**[1,2,±], **Sam de Blank**[1,2,±], **Mario Barrera Roman**[1,2], **Esmée J. van Vliet**[1,2], **Clara Martínez Mir**[1,2], **Hannah R. Johnson**[1,2], **Frank L. Bos**[1,2], **Raimond Heukers**[3], **Susana M. Chuva de Sousa Lopes**[4], **Jarno Drost**[1,2], **Johanna F. Dekkers**[1,2], **Ellen J. Wehrens**[1,2], **Anne C. Rios**[1,2,*]

[1]Princess Máxima Center for Pediatric Oncology, 3584 CS Utrecht, The Netherlands

[2]Cancer Genomics Netherlands, Oncode Institute, 3584 CG Utrecht, The Netherlands

[3]QVQ Holding BV, 3584 CL Utrecht, The Netherlands

[4]Department of Anatomy and Embryology, Leiden University Medical Center, 2333 ZC, Leiden, The Netherlands

Abstract

Despite advances in three-dimensional (3D) imaging, it remains challenging to profile all the cells within a large 3D tissue, including the morphology and organization of the many cell types present. Here, we introduce eight-color, multispectral, large-scale single-cell resolution 3D (mLSR-3D) imaging and image analysis software for the parallelized, deep learning-based segmentation of large numbers of single cells in tissues, called segmentation analysis by parallelization of 3D datasets (STAPL-3D). Applying the method to pediatric Wilms tumor, we extract molecular, spatial and morphological features of millions of cells and reconstruct the tumor's spatio-phenotypic patterning. *In situ* population profiling and pseudotime ordering reveals a highly disorganized spatial pattern in Wilms tumor compared to healthy fetal kidney, yet cellular profiles closely resembling human fetal kidney cells could be observed. In addition, we identify previously unreported tumor-specific populations, uniquely characterized by their spatial embedding or morphological attributes. Our results demonstrate the use of combining mLSR-3D and STAPL-3D to generate a comprehensive cellular map of human tumors.

[*]Correspondence to: a.c.rios@prinsesmaximacentrum.nl.
[#]These authors share first authorship
[±]These authors share second authorship

Author contributions

R.L.I. developed mLSR-3D and performed microscopy. M.K. developed STAPL-3D and performed the computational methods. R.L.I. and M.K. analyzed the data. S.B. assisted with computational analysis. C.M.M. assisted with microscopy. M.B.R. performed mLSR-3D imaging of breast and neuronal tumor tissue, rendered data and made movies. E.J.V. performed mLSR-3D imaging of breast and neuronal tumor tissue. M.A. performed quality control and computational analysis. J.F.D. provided organoid and xenograft material. H.R.J. assisted with sample preparation. F.L.B. provided microscopy support. R.H. provided the NCAM nanobody used in this study. S.M.C.S.L provided human fetal material. J.F.D., M.A., E.J.V., S.B., F.L.B and J.D.. provided critical feedback on the work. R.L.I., M.K. and A.C.R. designed the study and wrote the manuscript with support from E.J.W.. A.C.R. supervised this work.

Competing interests statement

We have no competing interests.

Europe PMC Funders Author Manuscripts

Single-cell resolution volumetric imaging permits the exploration of intact tissues[1–4], retaining spatial and geometrical information that is often lost through tissue dissociation in other single-cell technologies[5]. It thereby has the important advantage of revealing—in a single overview—the relationships between diverse cell types that both normal organ development and cellular function depend on, and how this is shifted under pathological conditions, such as cancer. Despite advances in 3D image processing, including nuclear and membrane segmentation methods[6–11] and large-scale nuclei counting of intact human organs[12], delineating the exact cellular organization of large human tissues at the single-cell level remains a challenge. The high number of cellular subsets and their various morphologies and configurations all complicate accurate single-cell identification and profiling. While challenging, such an approach would be highly informative, as it creates a single-cell readout that retains spatial and morphometric information and can thereby phenotype cells in the context of their native tissue environment. Therefore, to fully exploit the potential of volumetric imaging, we here developed multispectral Large-scale Single-cell Resolution 3D (mLSR-3D) imaging with ‘on-the-fly’ linear unmixing for single-scan acquisition of 8 spectrally-resolved fluorophores. Combined with SegmenTation Analysis by ParaLlelization of 3D Datasets (STAPL-3D), an automated pipeline for compartment-specific feature extraction, it enables *in situ* analysis of millions of cells in tissue (Fig. 1a, Supplementary Video 1).

Results

To interrogate the cellular biology and heterogeneity of tissues, we sought for an imaging strategy to timely image multiple markers in 3D (Supplementary Fig. 1). We first defined a combination of 8 fluorophores (out of 21 fluorophores compatible with linear unmixing of lambda-stacks[13] that we tested (Fig. 1b, Supplementary Fig. 1b)). Their reference emission spectra were used for accurate unmixing during single-scan acquisition (Supplementary Fig. 2) without the need for individual fluorophore control samples, a major advantage compared to recent methods relying on post-acquisition compensation, thereby generating additional data files[1,2,14]. When performing ‘on the fly’ spectral unmixing, equal signal detection is required, which is challenging for 8 fluorophores and cannot be achieved through adjusting laser power or detection settings. To overcome this issue, we developed a large-content intensity equalization assay for mLSR-3D-imaging to ensure balanced fluorescent intensities through the immunolabeling process. Using this assay, we tested over 60 antibodies and dyes for optimal 8-color staining (Supplementary Fig. 1, Supplementary Table 1) and selected 5 markers of interest based on recent scRNAseq data[15] to label a broad range of early nephrogenic structures of Human Fetal Kidney (HFK) development, as well as DAPI to stain nuclei, Phalloidin to label the F-actin network and KI67 to mark cycling cells (Fig. 1c-f). To facilitate the use of 8 fluorophores, we implemented a 5-day protocol, consisting of 3 rounds of labeling for flexible use of multiple species of primary antibodies combined with fluorescent secondaries, as well as direct conjugates, followed by a non-toxic clearing step with FUnGI[3] that preserves cell morphology and tissue architecture (Supplementary Fig. 3). This versatile protocol can be applied to a wide range of tissues, as demonstrated by 8-color mLSR-3D imaging of xenografted human organoid-derived breast tumors (Supplementary Fig. 4a), associated breast cancer organoids cultured *in vitro* (Supplementary Fig. 4b),

and biopsy-derived human central nervous system tumor material (Supplementary Fig. 5). Therefore, this method enables acquisition of large-scale, multi-dimensional 3D datasets with drastic reduction in overall acquisition time, photobleaching by repetitive illumination, and data preprocessing and storage requirements.

We then developed the STAPL-3D pipeline for single-cell feature extraction from large 3D imaging dataset (Fig. 1a). First, to optimize mLSR3D datasets for subsequent analysis, we implemented the STAPL-3D preprocessing module (Supplementary Fig. 6a). It includes a novel channel-specific shading correction (Supplementary Fig. 6b and d) and a 3D inhomogeneity correction developed for Magnetic Resonance Imaging[16] to reduce technical background variations (Supplementary Fig. 6c and d). Furthermore, for high autofluorescence, observed in the AF488 channel in the kidney, we employed machine learning to generate voxelwise probability map[17] for KI67, enabling accurate quantification of cycling cells (Supplementary Fig. 7, Supplementary Table 2). Next, the STAPL-3D segmentation module (Supplementary Fig. 8a) segments the dataset into individual cells and subdivides each cell into nucleus and membrane (Fig. 1g). STAPL-3D makes optimal use of mLSR-3D data by combining membrane and nucleus channels to generate seeds, followed by a two-step watershed procedure expanding the seed into the nucleus and then filling the cell to the membrane boundary (Supplementary Fig. 8b & 9). For scalable processing, we designed STAPL-3D to be compatible with high performance computing for complete segmentation in a couple of hours, by distributing the various analysis steps over volumes, channels and datablocks (Fig. 1h, Supplementary Fig. 10). Yet, STAPL-3D also runs efficiently on laboratory workstations. Splitting the dataset into blocks generates seams of partially segmented cells touching the block borders, either resulting in substantial data loss by excluding them[18], or introducing artefacts to these cells. Therefore, we developed a zipping module that identifies erroneous segments, resegments them, and merges the blocks back into a single seamless segmented volume (Supplementary Fig. 8c and 9d).

To achieve maximum utility of STAPL-3D, we provide the option to use state-of-the-art deep learning segmentation methods within the pipeline (STAPL-3D^{DL}), by integrating 3D-UNET[10] to predict membrane probability and StarDist[11] to predict individual nuclei (Supplementary Fig. 11). Because manual segmentation proved practically unfeasible for mLSR-3D datasets (with 80 hours of labor required for 569 cells—which was insufficient for model training), we also provide a STAPL-3D module to generate large training datasets (Supplementary Fig. 11a) by co-acquisition of mLSR-3D data at typical resolution (yielding the training data) and at very high resolution (yielding the training labels at the same location). Furthermore, using these labeled datasets, we offer a segmentation parameter tuning module that uses Bayesian optimization to automatically choose parameters that result in the best segmentation quality (Supplementary Table 3; STAPL-3D^{FT}). Segmentation accuracy of the modules was assessed by comparing to an extensive (N=14,717 cells) and diverse set of cells (from 10 different area's of the kidney), that was segmented with high fidelity from datasets co-acquired at high resolution and manually curated afterwards to yield a ground truth dataset. Dice overlap, precision, recall and F-scores were computed as accuracy metrics (Supplementary Fig. 12, Supplementary Table 3). We obtained the highest $F_{1.5}$-score of 0.81 (±0.012 SEM) for STAPL-3D^{DL} trained on co-acquired mLSR-3D datasets (Supplementary Fig. 12b)

followed by STAPL-3D^{FT}($F_{1.5}$=0.75±0.014), both demonstrating a significant increase in performance over the generic deep learning model ($F_{1.5}$=0.71±0.014) and the non-tuned STAPL-3D pipeline ($F_{1.5}$=0.72±0.017). Furthermore, comparing extracted morphological features with the ground truth, showed little morphological divergence. Overall average percentual increases and decreases with respect to the ground truth were 5.929% (±2.905) and 7.198% (±1.740) for STAPL-3D^{DL} trained and 7.821% (±2.768) and 8.912% (±2.220) for STAPL-3D^{FT} (Supplementary Fig. 12c). This analysis thus confirms that once trained on the appropriate data, STAPL-3D^{DL} increases segmentation accuracy. Nevertheless, STAPL-3D already offers a robust segmentation pipeline with a good performance.

STAPL-3D extracts molecular marker intensities, as well as spatial and 3D morphological properties per segmented cellular compartment. By default, features are computed for the cell and the membranal and nuclear subsegments. Moreover, we show that mLSR-3D and STAPL-3D pipelines can be adapted to extract features from the cytosolic and even mitochondrial compartment when using Airyscan 3D imaging (Supplementary Fig. 13). The division into cellular compartments can be exploited to define compound features, for example cell polarity, estimated from the centers of mass of cell and nucleus (Supplementary Fig. 8d). We can obtain a complete set of approximately 800 features extracted from full cell, nucleus and membrane segment (Supplementary Table 4), with the option to select the features most relevant for the particular application (Fig. 1i). Altogether, STAPL-3D offers a scalable, modular and tunable analysis framework for advanced image preprocessing and cellular compartment-specific segmentation, towards reliable 3D feature extraction and profiling of millions of cells within tissue.

To showcase the potential of our mLSR-3D and STAPL-3D framework, we next performed spatio-phenotypic patterning of Wilms Tumor (WT) – a pediatric kidney cancer (Fig. 2 and Supplementary Fig. 14). Prevalence of WT in early childhood has been related to corruption of fetal nephrogenesis and, indeed, these tumors present with aberrant fetal cells[19,20]. Therefore, we aim to elucidate the *in situ* developmental patterning of WT in relation to HFK. To create a spatio-phenotypic reference map, we defined 11 known cell populations in our HFK sample (gestational week 16), distributed over 3 components of the developing kidney: nephrogenesis, collecting system and interstitium[15]. This was achieved through 3D UMAP projection[21] and clustering of the 2.1M cells × 19 features data matrix (Fig. 2b, Supplementary Fig. 15c), assigning the resulting 20 clusters to a particular population, based on molecular markers, but also indispensably aided by morphological features and spatial location (Fig. 2c-d, Supplementary Fig. 15, 16a-b). Having captured the spatio-phenotypic single-cell landscape of HFK in a classifier, cell types could also be predicted for the 1.8M cells segmented from the WT sample. Backprojection of population identity into the dataset revealed a highly disorganized spatial pattern in WT compared to HFK (Fig. 2d-e), yet nephrogenic-like structures could be identified. These structures consisted predominantly of distal tubule (DT), ureteric bud (UB) and renal vesicle/S-shaped-body (RV/SSB) cells (Fig. 2f), and sporadically contained cell clusters of more undifferentiated cap mesenchyme (CM)-like cells (Fig. 2g). Indeed, in relation to HFK, the epithelial components of this particular WT were enlarged: 1.2 times for UB, 1.6 times for DT and, most notably, 2.8 times for early epithelial RV/SSB (Fig. 2h). Pseudotime ordering of nephrogenic cells revealed a single trajectory for early progenitors (CM) to committed progenitors (RV/SSB),

branching into three late populations (DT, proximal tubule progenitor/convoluted tubule (PTP/PCT) and glomerulus (GL)) (Fig. 2i). The position of the center of mass in this pseudotime UMAP showed a shift for the CM node towards the enlarged RV/SSB node in WT, indicative of a more committed progenitor fate for this cluster as compared to HFK. Inclusion of KI67 in our set of markers allowed us to provide insight into the mechanism underlying this developmental pattern (Fig. 2j). From the pseudotime-ordered cells, we could identify a peak of cycling cells during the RV/SSB stage in HFK. WT showed overall increased cycling compared to HFK; in particular high in RV/SSB, but also in their developmental progenitors (CM) (Fig. 2j). Thus, the enlarged RV/SSB cluster in the WT sample (Fig. 2h) likely results from both intrinsic cycling properties of this cluster, as well as a transitioning progenitor population from CM fueling this compartment. Hence, through profiling population distribution and pseudotime ordering, we could begin to untangle the *in situ* heterogeneity of WT in relation to its developmental origin.

To dive deeper into the spatio-phenotypic traits specific to WT, we next created a joint UMAP for HFK and WT and identified 6 cellular clusters that largely reflect conventional WT classification (epithelium, stroma and blastema[22]) (Fig. 3a-b). Based on their spatio-phenotypic features, we describe them as differentiated and undifferentiated epithelium, blastema, two stroma clusters, but also a small GL population (Fig. 3b, Supplementary Fig. 15). Although varying in contribution, all six clusters contained both healthy kidney and tumor cells, confirming the strong fetal resemblance of this tumor (Fig. 3c). Interestingly, cluster backprojection reveals one stroma-like compartment surrounding the epithelial/blastemal clusters (the surrounding stroma) while the other locates within these structures (the inner stroma) (Fig. 3d-e). This nicely demonstrates the usefulness of maintaining tissue-context to reveal differential spatial embedding of populations. The relative cluster sizes resulting from our classification (46.7% epithelium, 51% stroma and 2.3% blastema) (Supplementary Table 5), closely agree with histopathological scoring of the WT sample (50% epithelium, 45% stroma and 5% blastema), providing confidence in the obtained classification. Yet our approach goes beyond conventional classification and offers a more in-depth characterization (i.e. differentiated versus undifferentiated epithelium and two spatially resolved stromal-like compartments). In addition, we identified a tumor-specific population within the surrounding stroma through subclustering (Fig. 3f-g). Unexpectedly, cells belonging to subcluster 5 showed a high expression of SIX2, as compared to the remaining surrounding stroma (Fig. 3h), and are spindle-shaped (Fig. 3i), unlike conventional $SIX2^+$ round-shaped blastemal cells in WT (Fig. 3i-j; high-extent). This may be of particular significance, because SIX2 is involved in maintaining the undifferentiated and proliferative state of HFK CM and WT blastema cells[23], the latter known to be associated with poor prognosis if prevalent after chemotherapy[22]. 24% of WT cells of this subcluster 5 expressed SIX2 to a similar intensity as blastema cells (i.e. falling within or above the blastema interquartile range). Even though they represent only 2% of the entire WT sample, these SIX2-high stromal-like cells are present in substantial and thereby perhaps consequential amounts (36,242 cells). Although clinical importance of the identified cell profiles remains to be determined, we demonstrated that the combined application of mLSR-3D and STAPL-3D offers the potential to generate new insights into tumor

biology by accurate cell subset quantification and identification of novel spatio-phenotypic signatures.

Discussion

In sum, we here provide a targeted *in situ* profiling approach to exploit molecular, morphological and spatial features of millions of cells from 3D imaging data. In line with recent advances in multiplexed proteomics and spatial transcriptomics[1,3,4,24–26], we envision our single-cell technology a key step forward towards unravelling the complex cellular organization of organs and their associated tumors with particular promise for capturing essential spatio-phenotypic hallmarks of tumorigenesis.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We are grateful for the technical support from the Princess Máxima Center for Pediatric Oncology and Zeiss for imaging support. We would like to acknowledge the Gynaikon Clinic in Rotterdam for their efforts to provide the human fetal material and the laboratory of Hans Clevers and the Hubrecht Organoid Technology (HUB, www.hub4organoids.nl) for access to the Breast cancer organoid biobank. We also would like to acknowledge the Utrecht Bioinformatics Center High Performance Computing Facility for data processing infrastructure. We thank R.R. de Krijger for useful discussions. All the imaging was performed at the Princess Máxima Imaging Center. This work was financially supported by the Princess Máxima Center for Pediatric Oncology and St. Baldrick's Robert J. Arceci International Innovation award. J.F.D. is supported by a VENI grant from the Netherlands Organisation for Scientific Research (NWO). A.C.R is supported by an ERC-starting grant 2019 project 804412.

References

1. Coutu DL, Kokkaliaris KD, Kunz L, Schroeder T. Multicolor quantitative confocal imaging cytometry. Nat Methods. 2018; 15 :39–46. [PubMed: 29320487]
2. Li W, Germain RN, Gerner MY. High-dimensional cell-level analysis of tissues with Ce3D multiplex volume imaging. Nat Protoc. 2019; 14 :1708–1733. DOI: 10.1038/s41596-019-0156-4 [PubMed: 31028373]
3. Rios AC, et al. Intraclonal Plasticity in Mammary Tumors Revealed through Large-Scale Single-Cell Resolution 3D Imaging. Cancer Cell. 2019; 35 :618–632. e6 doi: 10.1016/j.ccell.2019.02.010 [PubMed: 30930118]
4. Segovia-Miranda F, et al. Three-dimensional spatially resolved geometrical and functional models of human liver tissue reveal new aspects of NAFLD progression. Nat Med. 2019; 25 :1885–1893. DOI: 10.1038/s41591-019-0660-7 [PubMed: 31792455]
5. Steinert EM, et al. Quantifying memory CD8 T cells reveals regionalization of immunosurveillance. Cell. 2015; 161 :737–749. DOI: 10.1016/j.cell.2015.03.031 [PubMed: 25957682]
6. Mosaliganti KR, Noche RR, Xiong F, Swinburne IA, Megason SG. ACME: Automated Cell Morphology Extractor for Comprehensive Reconstruction of Cell Membranes. PLoS Comput Biol. 2012; 8 doi: 10.1371/journal.pcbi.1002780
7. Stegmaier J, et al. Real-Time Three-Dimensional Cell Segmentation in Large-Scale Microscopy Data of Developing Embryos. Dev Cell. 2016; 36 :225–240. [PubMed: 26812020]
8. McQuin C, et al. CellProfiler 3.0: Next-generation image processing for biology. PLoS Biol. 2018; 16 doi: 10.1371/journal.pbio.2005970
9. Dunn KW, et al. DeepSynth: Three-dimensional nuclear segmentation of biological images using neural networks trained with synthetic data. Sci Rep. 2019; 9 doi: 10.1038/s41598-019-54244-5

10. Wolny A, et al. Accurate and versatile 3D segmentation of plant tissues at cellular resolution. Elife. 2020; 9 :1–34. DOI: 10.7554/eLife.57613
11. Weigert, M; Schmidt, U; Haase, R; Sugawara, K; Myers, G. Star-convex polyhedra for 3D object detection and segmentation in microscopy; Proceedings - 2020 IEEE Winter Conference on Applications of Computer Vision, WACV 2020; 2020.
12. Zhao S, et al. Cellular and Molecular Probing of Intact Human Organs. Cell. 2020; 180 :796–812. e19 [PubMed: 32059778]
13. Kraus B, Ziegler M, Wolff H. Linear fluorescence unmixing in cell biological research. Mod Res Educ Top Microsc. 2007 :863–872.
14. Valm AM, et al. Applying systems-level spectral imaging and analysis to reveal the organelle interactome. Nature. 2017; 546 :162–167. [PubMed: 28538724]
15. Hochane M, et al. Single-cell transcriptomics reveals gene expression dynamics of human fetal kidney development. PLoS Biol. 2019; 17 doi: 10.1371/journal.pbio.3000152
16. Tustison NJ, et al. N4ITK: Improved N3 bias correction. IEEE Trans Med Imaging. 2010; 29 :1310–1320. [PubMed: 20378467]
17. Berg S, et al. ilastik: interactive machine learning for (bio)image analysis. Nat Methods. 2019; 16 :1226–1232. [PubMed: 31570887]
18. Gut G, Herrmann MD, Pelkmans L. Multiplexed protein maps link subcellular organization to cellular states. Science (80-.). 2018; 361 doi: 10.1126/science.aar7042
19. Young MD, et al. Single-cell transcriptomes from human kidneys reveal the cellular identity of renal tumors. Science (80-.). 2018; 361 :594–599. DOI: 10.1126/science.aat1699
20. Young MD, et al. Single cell derived mRNA signals across human kidney tumors. bioRxiv. 2020
21. McInnes, L. PCA, t-SNE, and UMAP: Modern Approaches to Dimension Reduction; PyData Conference 2018; 2018.
22. Reinhard H, et al. Outcome of relapses of nephroblastoma in patients registered in the SIOP/GPOH trials and studies. Oncol Rep. 2008; 20 :463–467. DOI: 10.3892/or_00000029 [PubMed: 18636212]
23. Wegert J, et al. Mutations in the SIX1/2 Pathway and the DROSHA/DGCR8 miRNA Microprocessor Complex Underlie High-Risk Blastemal Type Wilms Tumors. Cancer Cell. 2015; 27 :298–311. [PubMed: 25670083]
24. Glaser AK, et al. Multi-immersion open-top light-sheet microscope for high-throughput imaging of cleared tissues. Nat Commun. 2019; 10
25. Stoltzfus CR, et al. CytoMAP: A Spatial Analysis Toolbox Reveals Features of Myeloid Cell Organization in Lymphoid Tissues. Cell Rep. 2020; 31
26. Merritt CR, et al. Multiplex digital spatial profiling of proteins and RNA in fixed tissue. Nat Biotechnol. 2020; 38 :586–599. DOI: 10.1038/s41587-020-0472-9 [PubMed: 32393914]

Europe PMC Funders Author Manuscripts

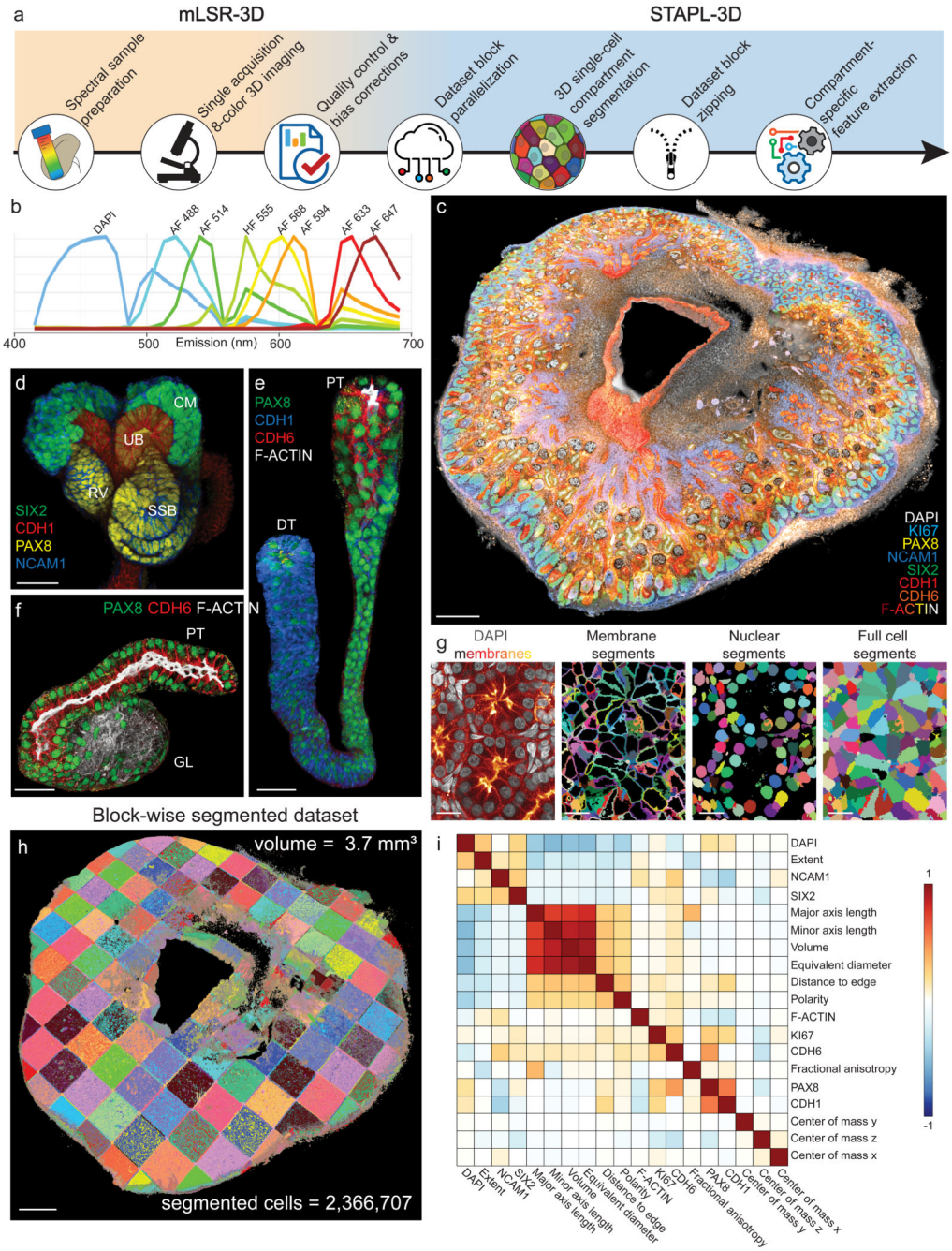

Figure 1. mLSR-3D imaging and STAPL-3D.

(**a**) Schematic overview of mLSR-3D and STAPL-3D (**b**) Normalized emission reference spectra. (**c**) mLSR-3D visualization of HFK (16 weeks of gestation) labeled for DAPI (grey), KI67 (cyan), PAX8 (yellow), NCAM1 (blue), SIX2 (green), CDH1 (red), CDH6 (orange) and F-ACTIN (gradient: red-yellow-white). Scale bar 500 μm. (**d-f**) 3D zooms of masked nephrogenic structures. (**d**) CM = cap mesenchyme, RV = renal vesicle, SSB = S-shaped body, UB = ureteric bud. SIX2 (green), CDH1 (red), PAX8 (yellow), NCAM1 (blue). (**e**) Loop of Henle with proximal tubule (PT) connecting to the distal tubule (DT).

Europe PMC Funders Author Manuscripts

PAX8 (green), CDH1 (blue), CDH6 (red), F-ACTIN (grey). (**f**) Proximal tubule (PT) connecting to the glomerulus (GL). PAX8 (green), CDH6 (red), F-ACTIN (grey). Scale bars 50 μm. (**g**) Optical section demonstrating cell compartment segmentation with STAPL-3D. DAPI (grey) and weighted mean of all membrane channels (red-yellow-white gradient). Segments are randomly colored. Scale bar 20 μm. (**h**) Volumetric rendering of the block-wise segmentation. Number of blocks = 182. Scale bar 500 μm. These experiments (c-h) were perfomed independently at least 4 times with similar results (Supplementary Fig. 10) (**i**) Pearson correlation heatmap of selected features, reordered by hierarchical clustering.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

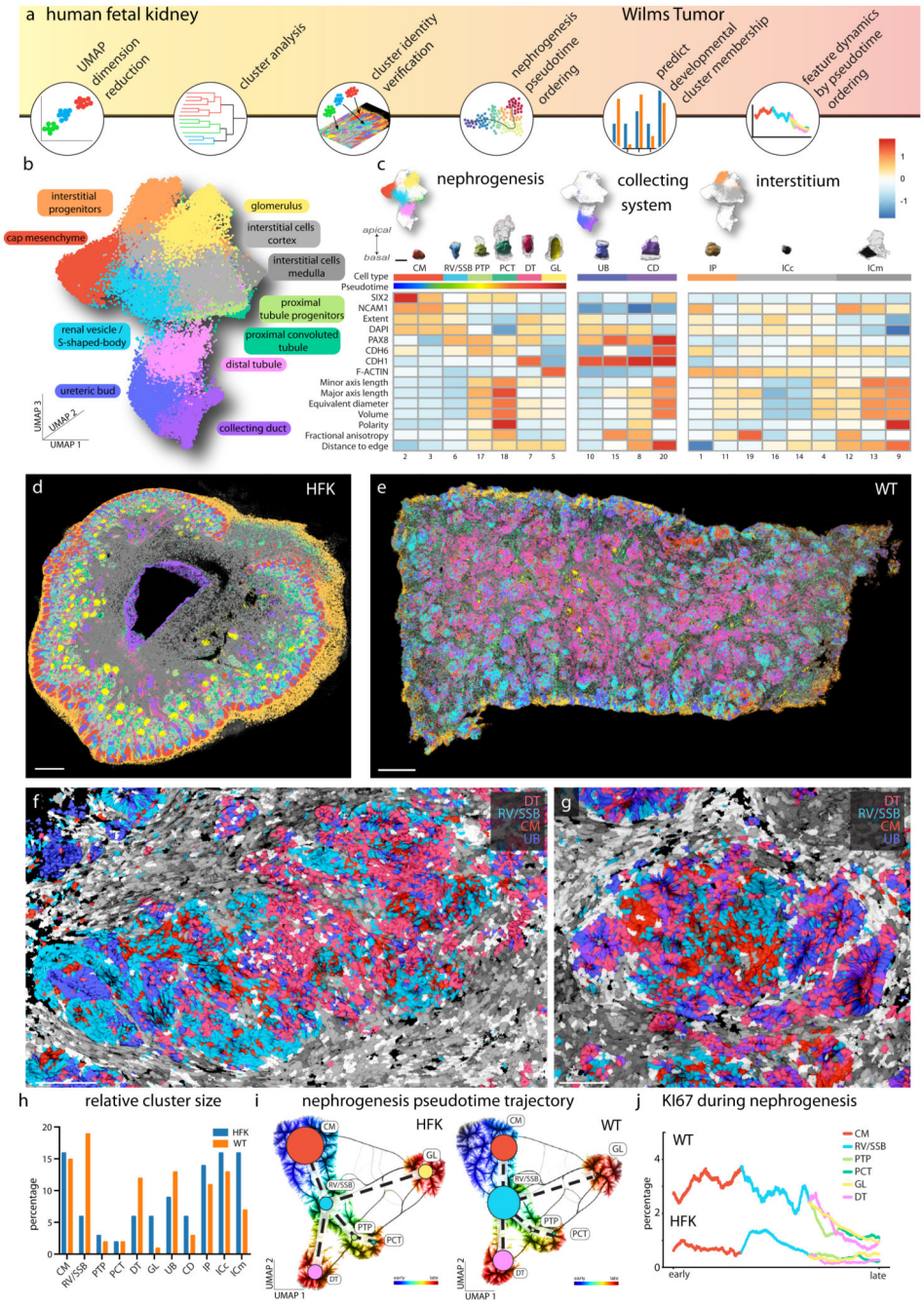

Figure 2. Spatio-phenotypic patterning of HFK and WT reveals expanded, cycling, early epithelial compartment in WT.

(**a**) Schematic overview of the analysis strategy. (**b**) 3D UMAP rendered for 50,000 HFK cells. Colors and labels correspond to cell identity. (**c**) Heatmap of log-scaled median feature values (blue-white-red gradient) per identified cluster, subdivided into components of the HFK: nephrogenesis (CM = cap mesenchyme, RV/SSB = renal vesicle/S-shaped body, PTP = proximal tubule progenitor, PCT = proximal convoluted tubule, DT = distal tubule, GL = glomerulus), collecting system (UB = ureteric bud, CD = collecting duct) and

Europe PMC Funders Author Manuscripts

interstitium (IP = interstitial progenitors, ICc = interstitial cells cortex, ICm = interstitial cells medulla). Clusters are numbered from 1 to 20 according to descending cluster size. Typical 3D segmented examples of each cell type are displayed above the heatmap, oriented apical to basal. Scale bar 10 μm. (**d-e**) 3D backprojection of the HFK and WT showing all single-cell segments colored for cell type identity. Scale bars 500 μm. (**f-g**) Optical sections showing backprojected cell types (DT, RV/SSB, CM and UB) of representative WT regions. Scale bars 70 μm. (**h**) Bar graph depicting relative cluster sizes per dataset in percentages. (**i**) UMAP depicting pseudotime ordering of HFK (left panel) and WT (right panel) cells belonging to nephrogenic clusters. Circle sizes correspond to the number of cells within each cluster and pseudotime is depicted by the rainbow gradient (early: blue – red: late). (**j**) KI67 positive fraction for WT (top) and HFK (bottom) plotted along the pseudotime trajectory of nephrogenic development. Line colors correspond to cell identities.

Europe PMC Funders Author Manuscripts

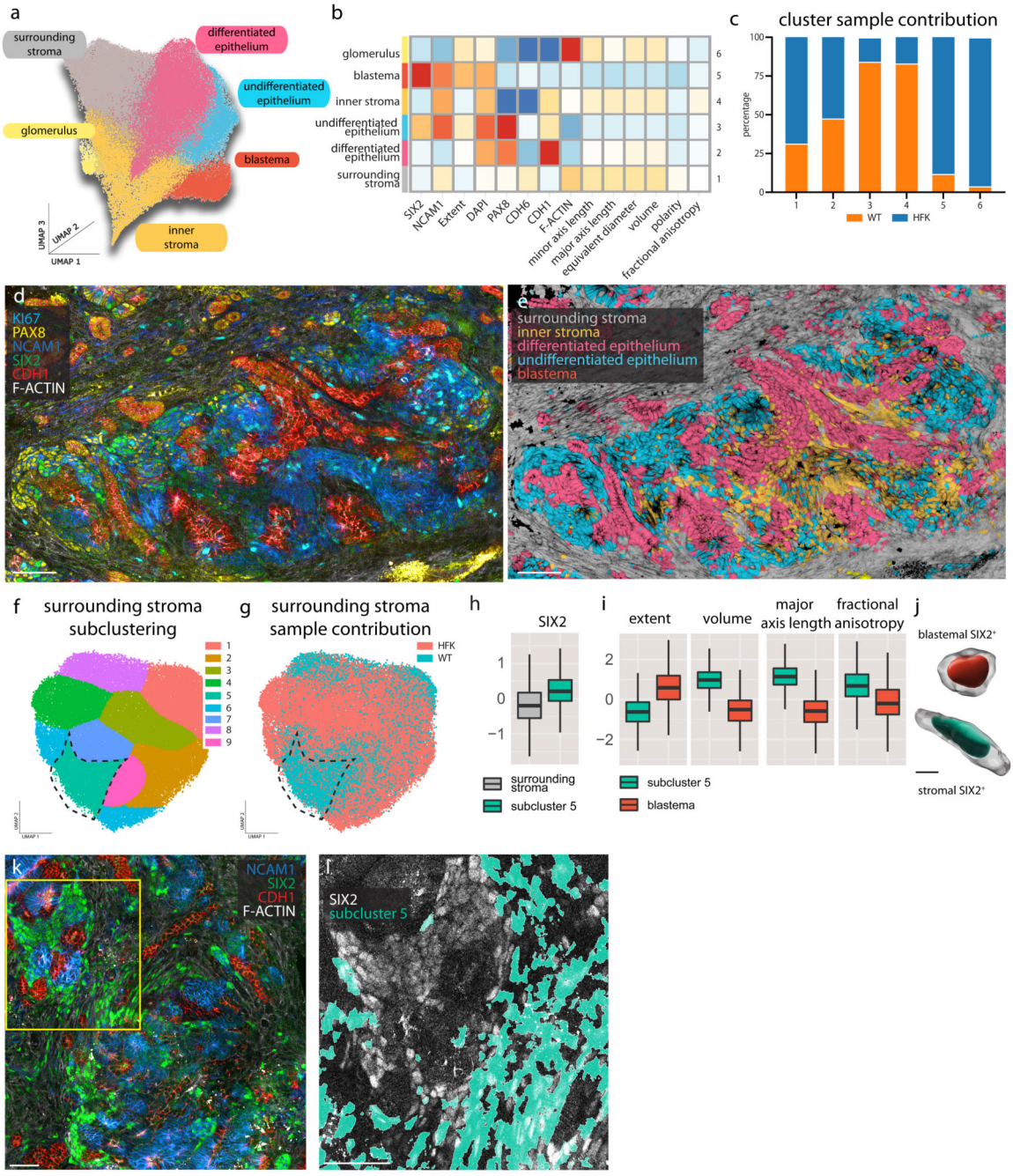

Figure 3. Comparative analysis reveals spindle shaped SIX2$^+$ cells in surrounding tumor stoma.
(**a**) Joint 3D UMAP rendered for 50,000 HFK and WT cells. Colors and labels correspond to cluster identity. (**b**) Heatmap of log-scaled median feature values (blue-white-red gradient) per identified cluster: clusters are numbered 1-6 according to descending cluster size. (**c**) Bar graph depicting relative contribution of each sample per cluster in percentages. (**d-e**) Optical sections showing fluorescent markers KI67 (cyan), PAX8 (yellow), NCAM1 (blue), SIX2 (green), CDH1 (red), and F-ACTIN (grey) (**d**) and backprojected cluster identities (**e**) of a representative WT region. Scale bars 70 μm. (**f-g**) UMAP rendered for 50,000 surrounding

stroma cells. Colors and labels correspond to (**f**) sub-cluster identity or (**g**) sample identity: blue = WT; red = HFK. (**h**) Boxplots showing log-scaled median SIX2 value of sub-cluster '5' (n = 194,218 cells) compared to the surrounding stroma main cluster (n = 1,650,651 cells). Centre: median, bounds: Q1-Q3, whiskers extend to minimum/maximum limited to 1.5 times the IQR (**i**) Boxplots showing log-scaled median values for extent, volume, major axis length and fractional anisotropy of sub-cluster '5' (n = 194,218 cells) compared to the blastema main cluster (n = 342,579 cells). Centre: median, bounds: Q1-Q3, whiskers extend to minimum/maximum limited to 1.5 times the IQR (**j**) Typical 3D segmented examples of a $SIX2^+$ blastemal (left) and $SIX2^+$ stromal cell (right). Scale bar 10 μm. (**k**) Optical section showing fluorescent markers NCAM1 (blue), SIX2 (green), CDH1 (red), F-ACTIN (grey) of a representative WT region. Scale bar 50 μm. (**l**) Magnification of indicated yellow area with back-projected subcluster '5' identities and SIX2 (grey). Scale bar 50 μm.

Europe PMC Funders Author Manuscripts