Single-cell transcriptomics identifies a distinct luminal progenitor cell type in distal prostate invagination tips

Wangxin Guo, Lin Li, Juan He, Zhuang Liu, Ming Han, Fei Li, Xinyi Xia, Xiaoyu Zhang, Yao Zhu, Yu Wei, Yunguang Li, Rebiguli Aji, Hao Dai, Hui Wei, Chunfeng Li, Yu Chen, Luonan Chen, and Dong Gao

INVENTORY

Supplemental Information contains the Supplemental Data (including 10 Extended Date Figures, 15 Supplementary Figures, 8 Supplementary Tables, and 2 Supplementary Notes).

Extended Data Figure 1-3 and 6 are related to Figure 1
Extended Data Figure 4 and 5 are related to Figure 2.
Extended Data Figure 7 is related to Figure 3.
Extended Data Figure 8 is related to Figure 5.
Extended Data Figure 9 and 10 are related to Figure 6.

Supplementary Figure 1-5 are related to Figure 1.
Supplementary Figure 6 and 7 are related to Figure 2.
Supplementary Figure 8-11 are related to Figure 3.
Supplementary Figure 12-15 are related to Figure 6.

Supplementary Table 1 (related to Figure 1). Cell type-specific signatures of mouse prostate.
Supplementary Table 2 (related to Figure 1). Go enrichment of Luminal-A, Luminal-B and Luminal-C cell population.
Supplementary Table 3 (related to Extended Figure 2). Cell type-specific signatures of VP, DLP, AP mouse prostate.
Supplementary Table 4 (related to Figure 2). Signatures of bulk Dist-Luminal-C cells, Prox-Luminal-C cells, Tacstd2-negative luminal cells and basal cells.
Supplementary Table 5 (related to Figure 6). Cell type-specific signatures of human prostate.
Supplementary Table 6 (related to Figure 6). Shared and different marker genes of the luminal-c cell clusters between human and mouse prostate cell clusters.
Supplementary Table 7. Primers and sgRNA used in this study.
Supplementary Table 8. Antibodies used in this study.

Supplementary Notes #1

Trajectory analysis of single cells. We used an R package Monocle2[1] (version 2.6.1) to construct pseudo-time trajectory of luminal cells (Luminal A, Luminal B and Luminal-C). We randomly selected 3,032 cells from 6,063 cells in proportion to the three cell types using an R function sample. Genes were selected if they were expressed in more than 10 cells, and their minimum mean expression value was 0.5. Differentially expressed genes of clusters were identified using differentialGeneTest with q-value < 0.01. Then, those genes were used to conduct the trajectory analysis. Cells were ordered along the trajectory which was visualized in a two-dimensional space. Monocle2 was used to perform cell trajectory analysis of luminal cells (Luminal-A, Luminal-B and Luminal-C).Pseudotime analysis demonstrated a diversion of Luminal-A and Luminal-B cells from Luminal-C cells (Supplementary Fig. 4a,b). Progenitor cells showed high expression levels of cell fate- or cell proliferation-associated genes, such as Sox9 and Psca (Supplementary Fig. 4c,d).

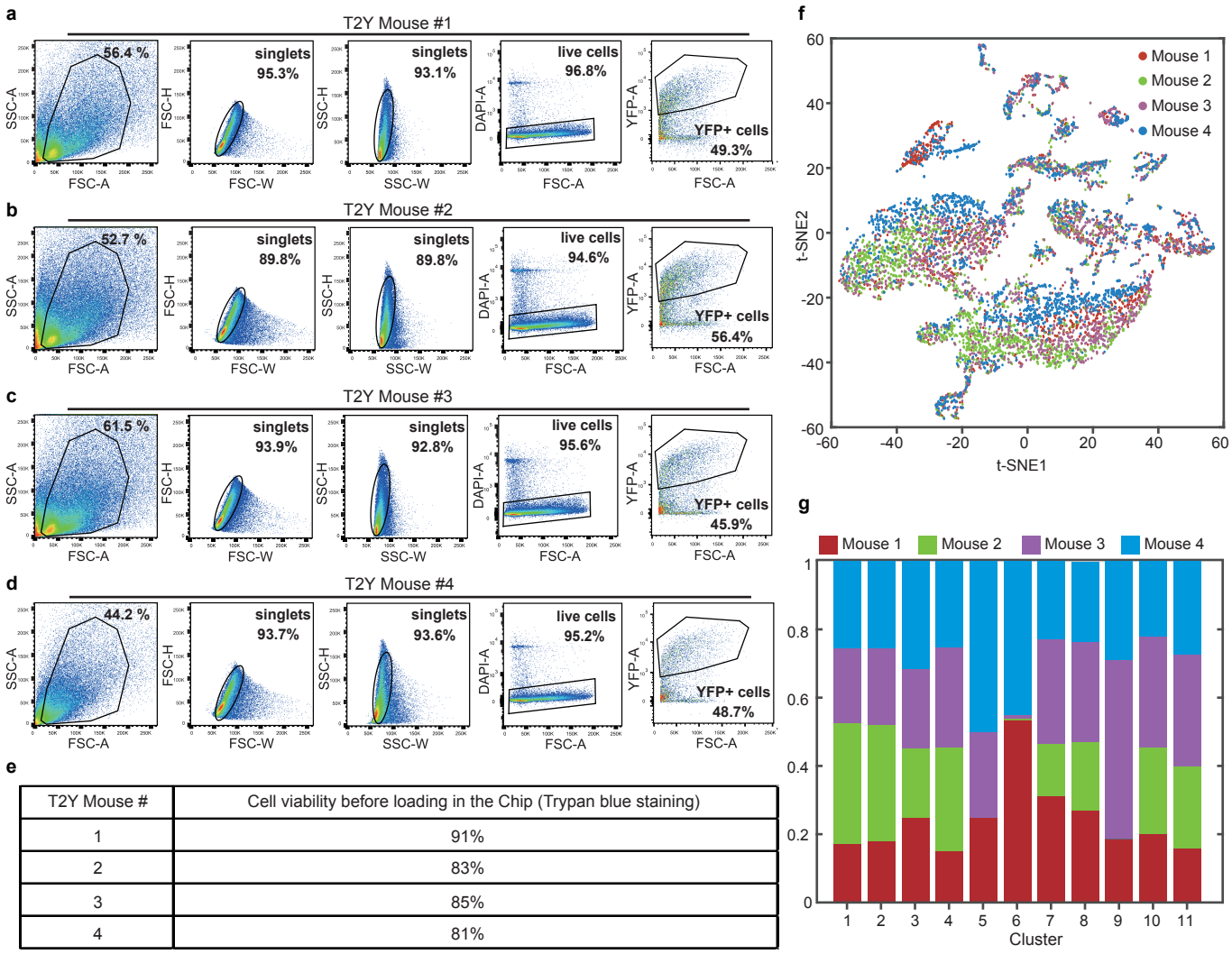
Supplementary Notes #2

To further confirm the existence of these luminal cell lineages, we profiled 4,100 individual cells from the freshly dissociated whole prostate of 1 wild-type (WT) mouse aged 10 weeks. Unbiased clustering analysis identified 5 distinct cell clusters (Supplementary Fig. 3a). To define the cellular identity of each mouse prostate cell cluster, we identified cluster-specific marker genes (Supplementary Fig. 3b-e). Notably, we found similar patterns for Luminal-A, Luminal-B and Luminal-C cells in the WT mouse prostate (Supplementary Fig. 3b-e). These results confirmed the existence of Luminal-A, Luminal-B and Luminal-C cell populations in the mouse prostate.

Supplementary Reference

1. Qiu, X. *et al.* Reversed graph embedding resolves complex single-cell trajectories. *Nat Methods* **14**, 979-982 (2017).

Supplementary Figure 1

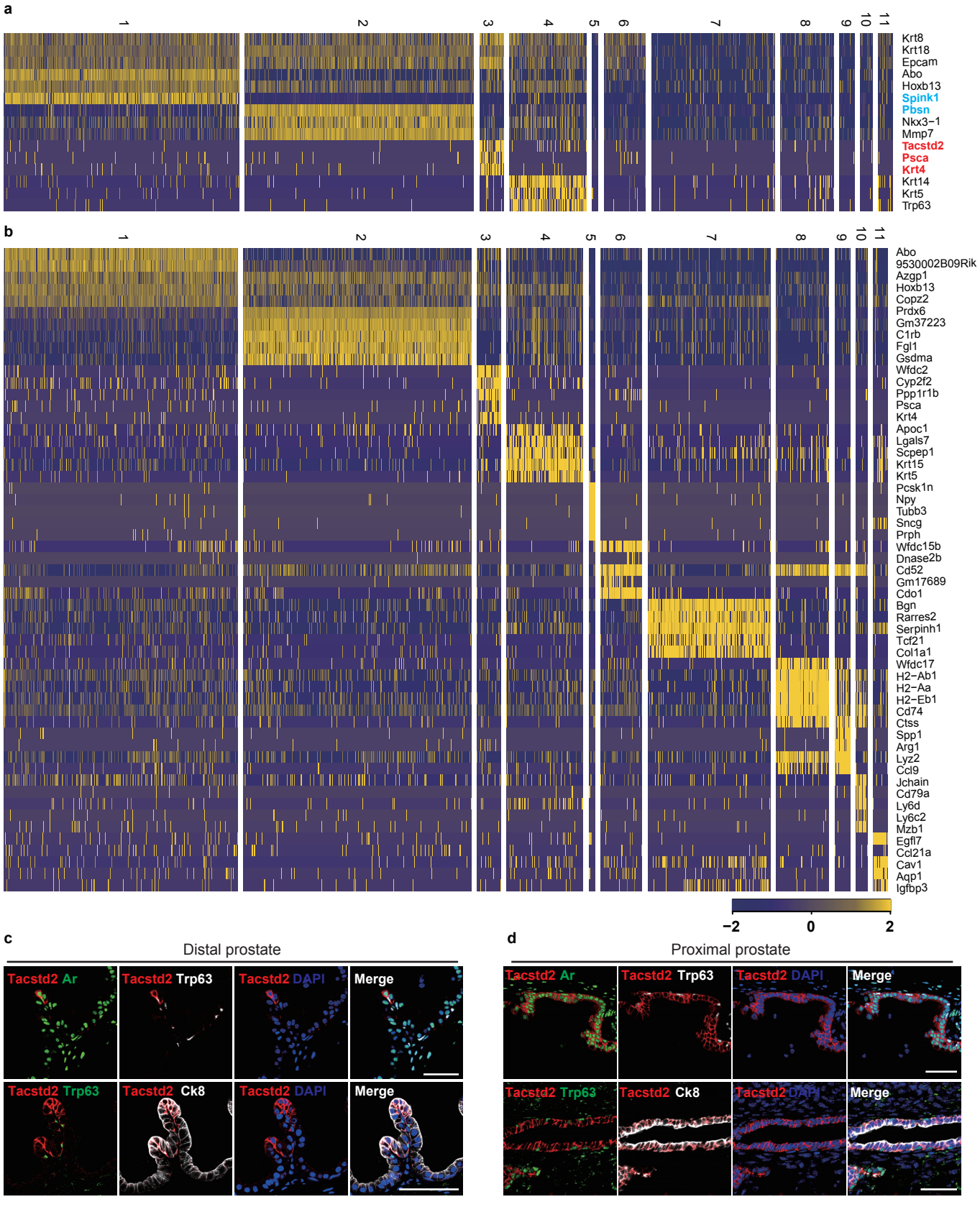


e

T2Y Mouse #	Cell viability before loading in the Chip (Trypan blue staining)
1	91%
2	83%
3	85%
4	81%

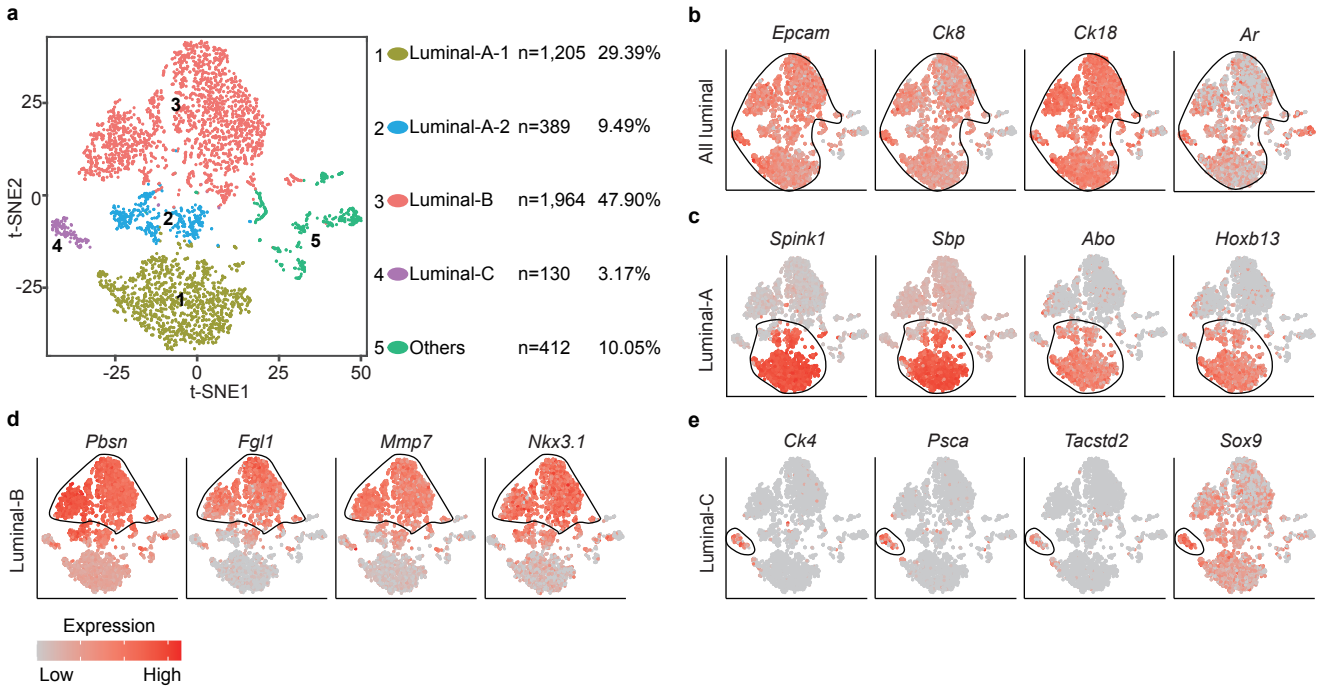
Supplementary Fig. 1 | Whole prostate cell sorting and distribution of cells from each mouse. a-d, FACS plots, gating strategy and cell viability of the 4 T2Y mouse prostates. FACS plots and gating strategy of the 4 T2Y mouse prostates. **e,** Cell viability of the sorted 4 T2Y mouse prostate cells before loading in the Single-Cell-A-Chip. **f,** T-SNE map shows the cell distribution from four different mouse prostates (n = 8,545 cells). **g,** Bar graph shows the percentage of the four different mice. 1-11 represents 11 different cell clusters and the colors represent each mouse.

Supplementary Figure 2



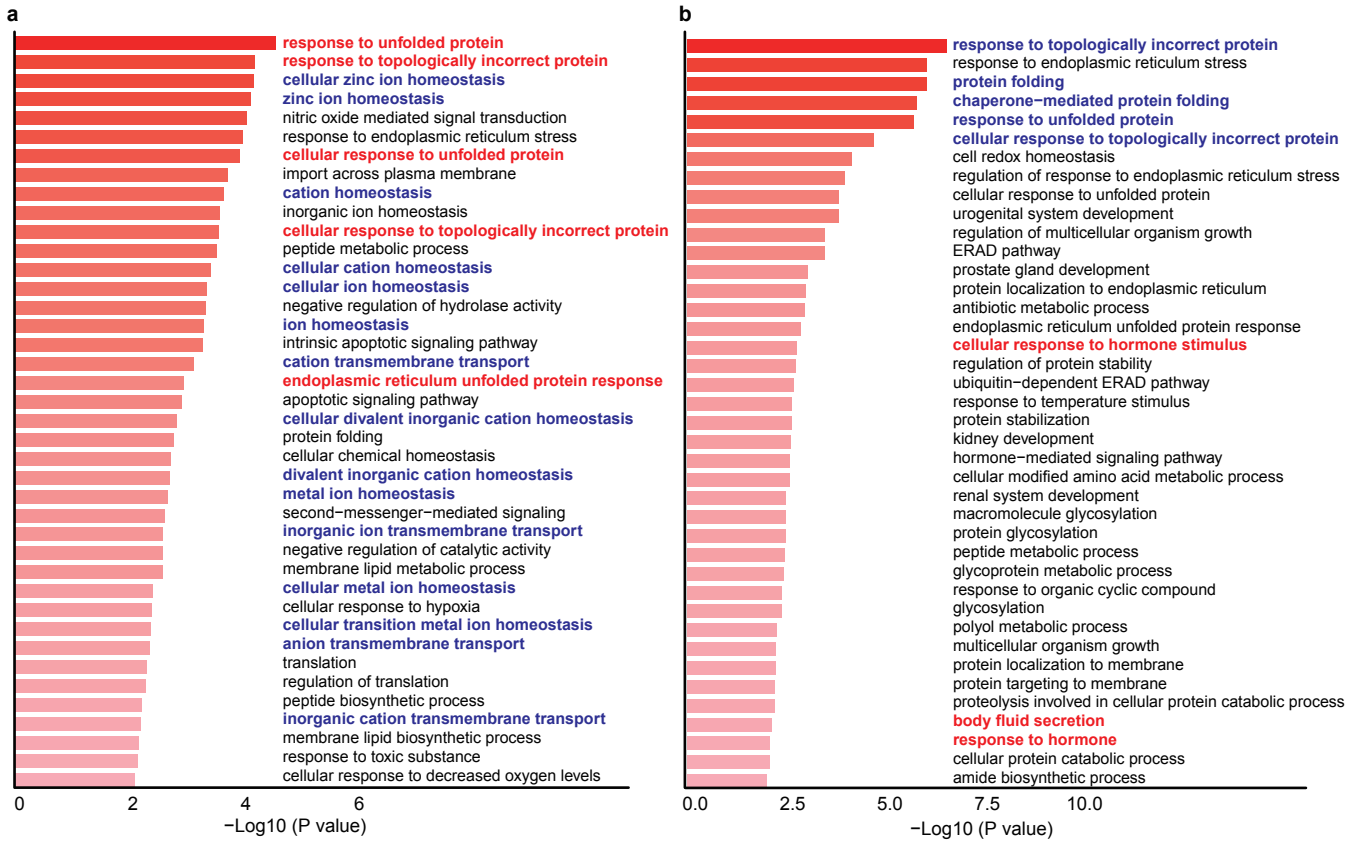
Supplementary Fig. 2 | Specific gene expression patterns for each cell type. a, Heatmap shows the relative expression level (row Z-score of log scale normalized read count) of prostate cells marker genes (rows) of each cluster across cells (columns, n=8,545 cells). Epcam for all the prostate epithelial cells; Ck8 and Ck18 for all the luminal cells; Abo, Hoxb13 and Spink1 for Luminal-A cells; Pbsn, Nkx3.1 and Mmp7 for Luminal-B cells; Tacstd2, Psca and Ck4 for Luminal-C cells; Ck5, Ck14 and Trp63 for Basal cells. **b,** Heatmap shows the relative expression level (row Z-score of log scale normalized read count) of top 5 marker genes (rows) of each cluster across cells (columns, n=8,545 cells). Luminal-C cells are AR and Ck8 double positive luminal cells. **c,d,** Immunostaining of Tacstd2, Ar, Trp63 and Ck8 shows Luminal-C cells express luminal cell specific marker Ck8 and Ar in distal prostate regions (**c**) and proximal prostate regions (**d**). 3 independent mice were used for the experiment with 4 replicates (**c, d**). 3 independent mice were used for the experiment with 4 replicates (**c, d**). Scale bars, 50 μm (**c,d**).

Supplementary Figure 3



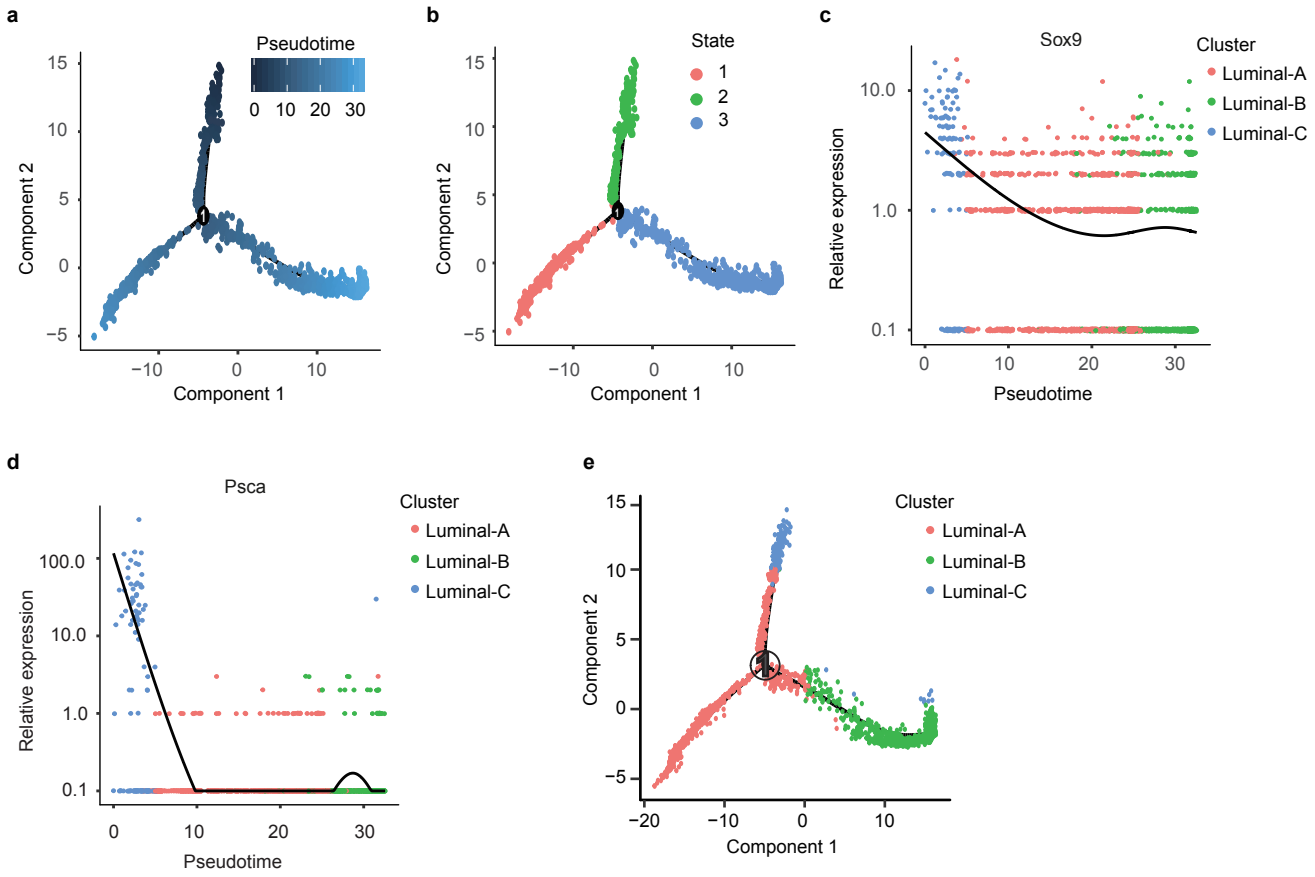
Supplementary Fig. 3 | Luminal cell populations in WT mouse prostate. a, Visualization of clustering of 4,100 WT mouse prostate single cells, based on the expression of known marker genes by t-SNE. **b-g,** T-SNE maps show the expression levels of marker genes across 5 clusters. Black circle indicate the Luminal clusters (n = 3,688 cells) (**b**), Luminal-A cluster (n = 1,594 cells) (**c**), Luminal-B cluster (n = 1,964 cells) (**d**), Luminal-C cluster (n = 130 cells) (**e**). T-SNE map shows cells that are colored by the log-scale normalized read count of marker genes.

Supplementary Figure 4



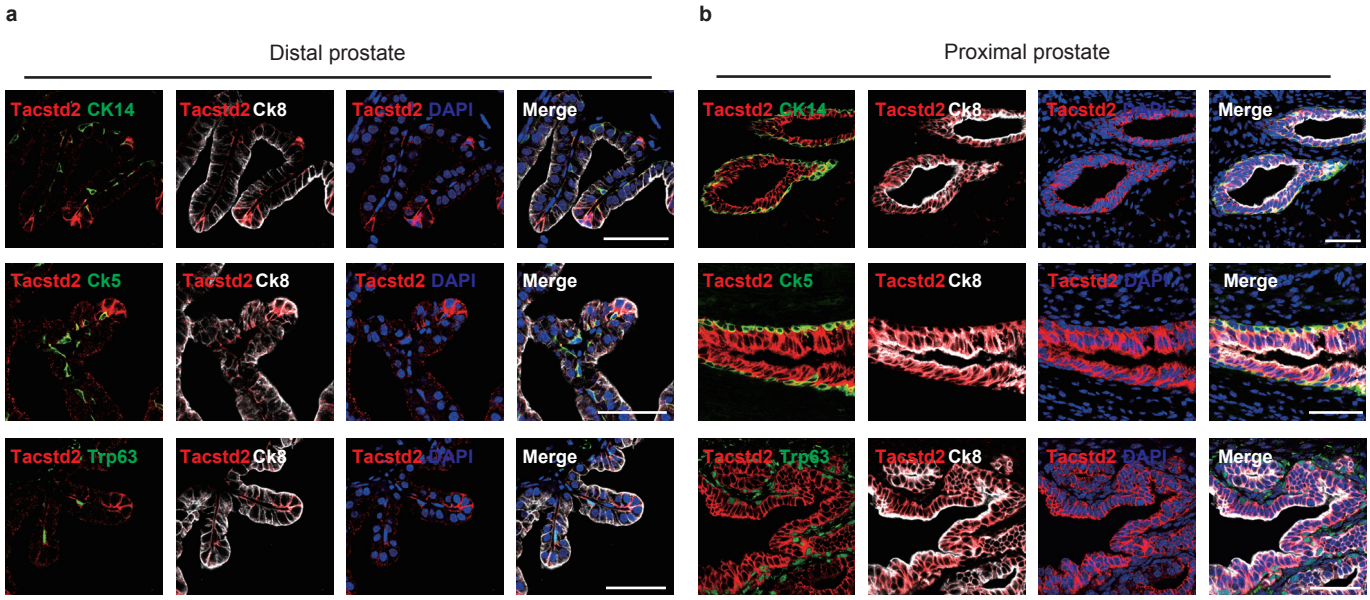
Supplementary Fig. 4 | Signature of Luminal-A and Luminal-B cell subtype are analyzed by GO enrichment assay. a, Luminal-A signature genes are enriched (n= 2,773 cells). Shown are the top 40 significantly enriched (Methods; P value < 0.01; -log10 (P value)) gene ontology terms in the gene signature for the Luminal-A subtype. **b**, Luminal-B signature genes are enriched (n = 2,233 cells). Shown are the top 40 significantly enriched (Methods; P value < 0.01; -log10 (P value)) gene ontology terms in the gene signature for the Luminal-B subtype.

Supplementary Figure 5



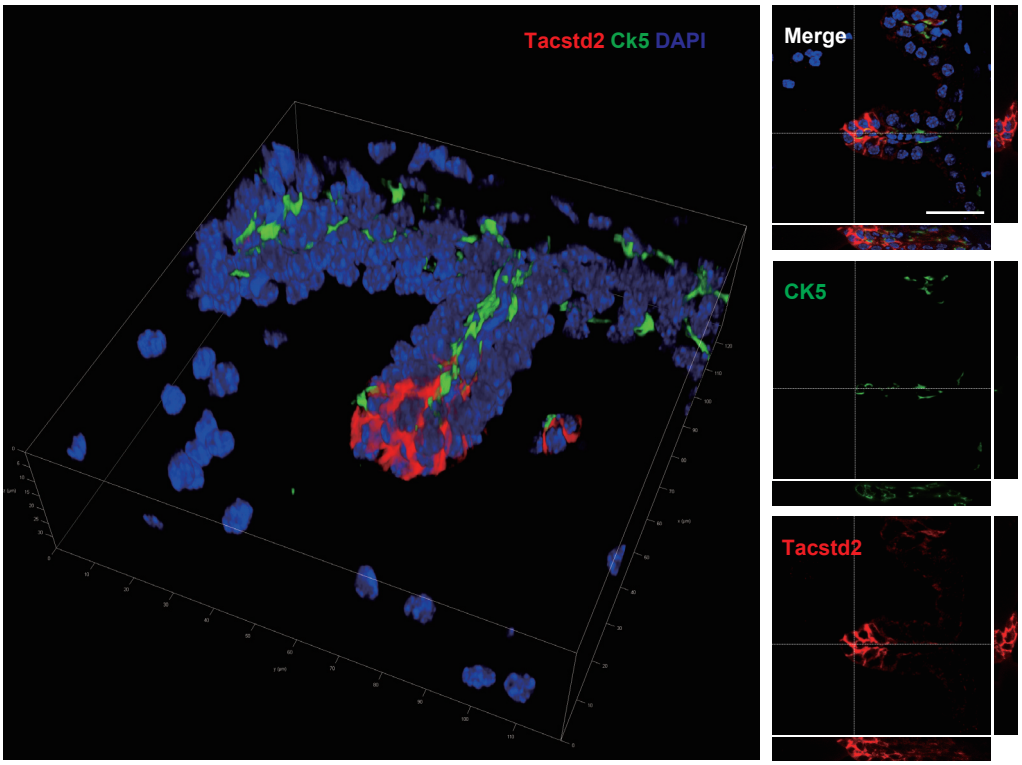
Supplementary Fig. 5 | Trajectory of luminal single cells conducted by monocle (n =3,032 cells). **a,** Trajectory map shows the pseudo-time of three types of luminal cells. **b,** The cell trajectory map shows three states which were identified based on their distribution. **c,d,** Relative expression level of cell-fate or cell-proliferation associated genes, such as Sox9 (c) and Psca (d) in three luminal cell clusters along the pseudo-time are shown. The line shows gene expression trend. **e,** Cell trajectory analysis of luminal cells (Luminal-A, Luminal-B and Luminal-C). Three clusters of luminal cells (1,032 luminal-A cells, 875 Luminal-B cells and 114 Luminal-C cells) were used for ordering and visualized in two-dimensional space with different colors.

Supplementary Figure 6



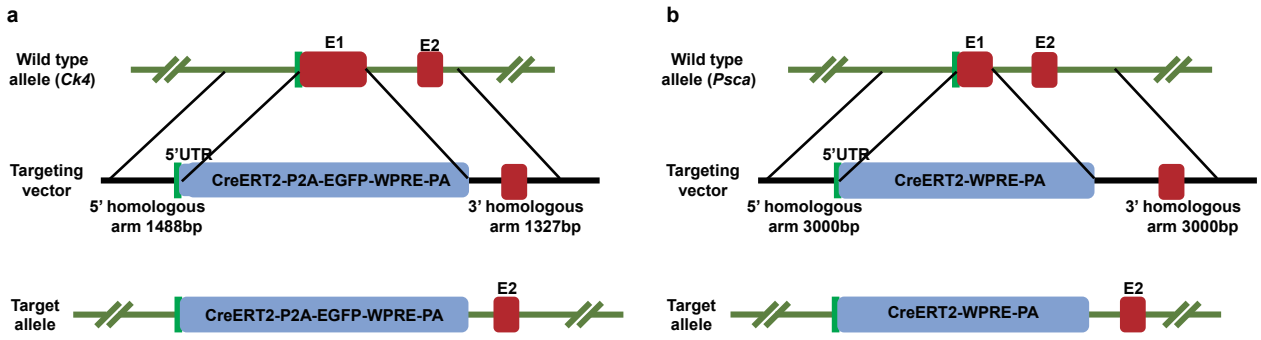
Supplementary Fig. 6 | Defining the location of Dist-Luminal-C and Prox-Luminal-C cells in mouse prostate. Dist-Luminal-C cells locate at invagination tips of prostate distal regions, while Luminal-A and Luminal-B clusters are interspersed distribute in the distal prostate. Prox-Luminal-C cells locate at the prostate proximal region. **a**,**b**, Co-immunofluorescence of Ck14, Ck5, Trp63, Ck8 with Tacstd2 in prostate distal regions (**a**) and proximal regions (**b**). 3 independent mice were used for the experiment with 4 replicates (**a,b**). Scale bars, 50 μm (**a**,**b**).

Supplementary Figure 7



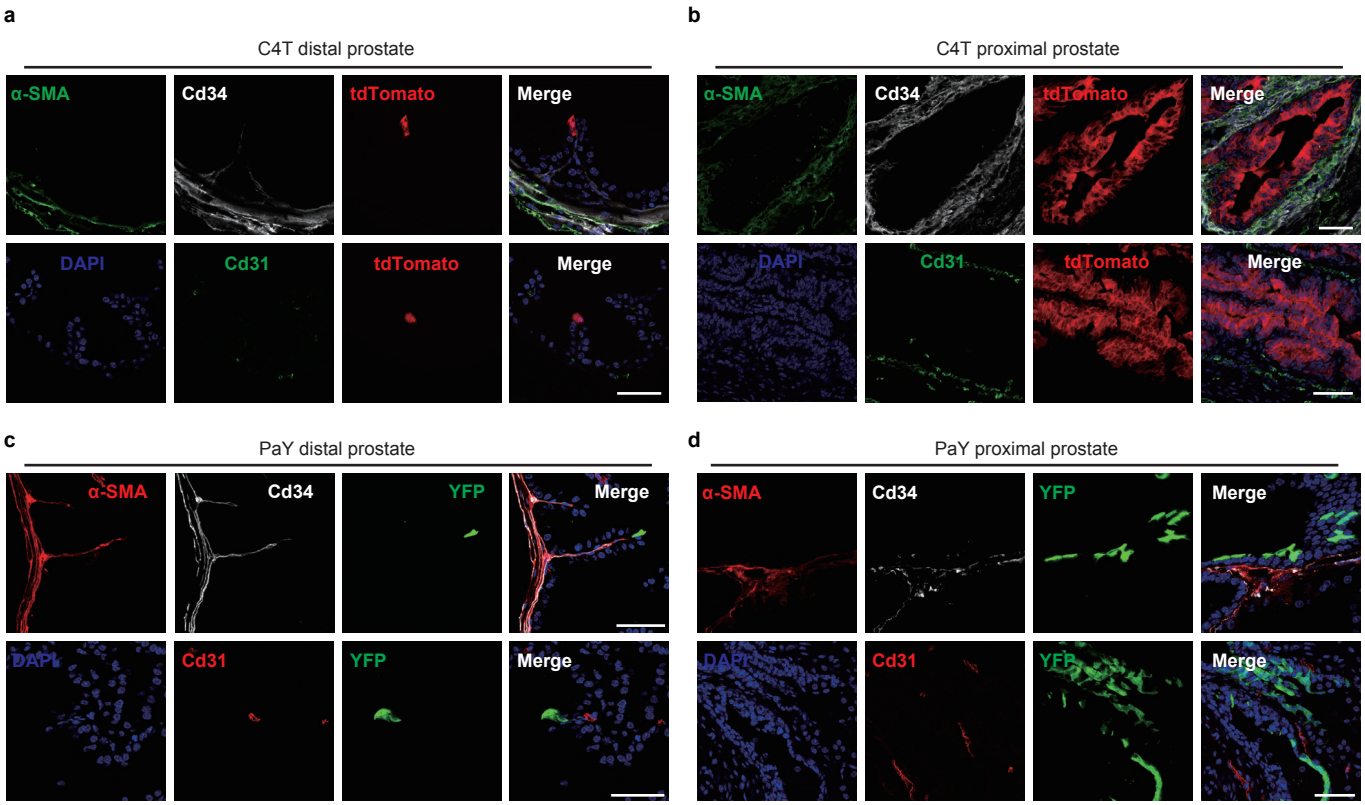
Supplementary Fig. 7 | Dist-Luminal-C cells reside in the invagination tips of distal mouse prostate. Co-immunofluorescence of Ck5 and Tacstd2 in thick slide of WT mouse distal prostate. These experiments were administered five times with similar results. Scale bars, 50 μm.

Supplementary Figure 8



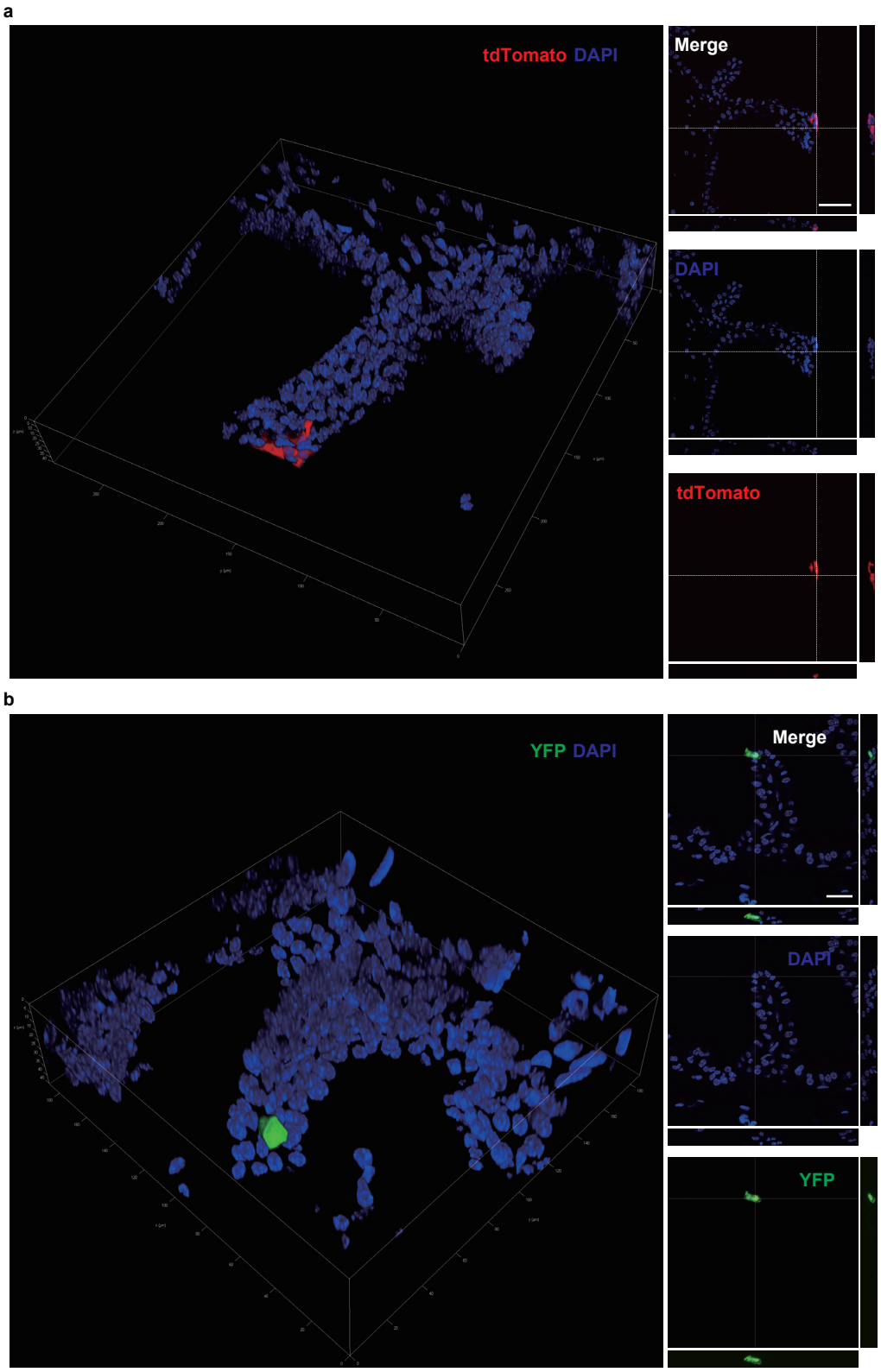
Supplementary Fig. 8 | The generation of Luminal-C-cell-lineage-specific mice through knock-in a CreERT2 cassette which expression was controlled by *Ck4* or *Psca* promoter. The strategies of the generation of $Ck4^{CreERT/+2}$ (**a**) and $Psca^{CreERT2/+}$ mice (**b**).

Supplementary Figure 9



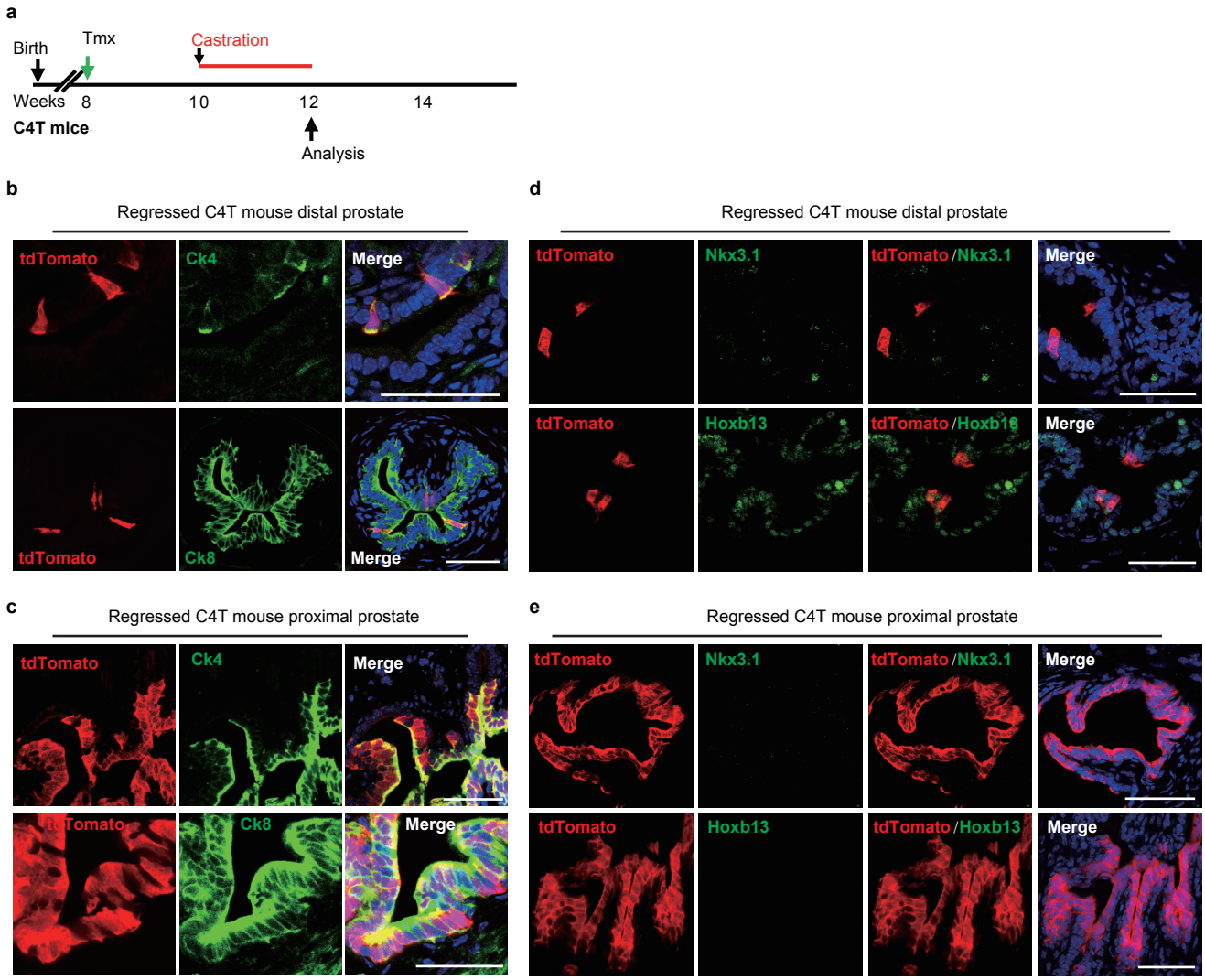
Supplementary Fig. 9 | Luminal-C cells are specially labelled in C4T and PaY mouse prostate a,**b**, Co-immunofluorescence of α-SMA, Cd34, Cd31, with endogenous tdTomato in prostate distal (**a**) and proximal regions (**b**) of C4T mice. **c,d,** Co-immunofluorescence of α-SMA, Cd34, Cd31, with endogenous YFP in prostate distal regions (**c**) and proximal regions (**d**) of PaY mice. 3 independent mice were used for each experiment. Scale bars, 50 μm (**a-d**).

Supplementary Figure 10



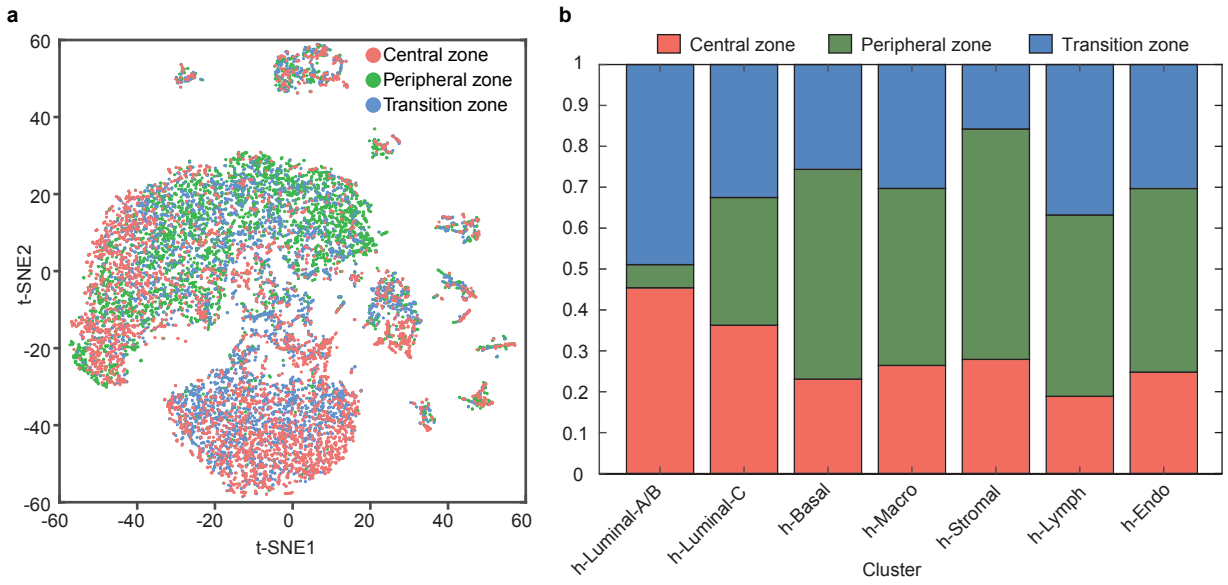
Supplementary Fig. 10 | Dist-Luminal-C cells reside in the invagination tips of C4T and PaY mouse prostate distal regions. a,b, 3D imaging for the prostate invagination tips in C4T (**a**) and PaY (**b**) mouse prostates. Scale bars, 50 μm.

Supplementary Figure 11



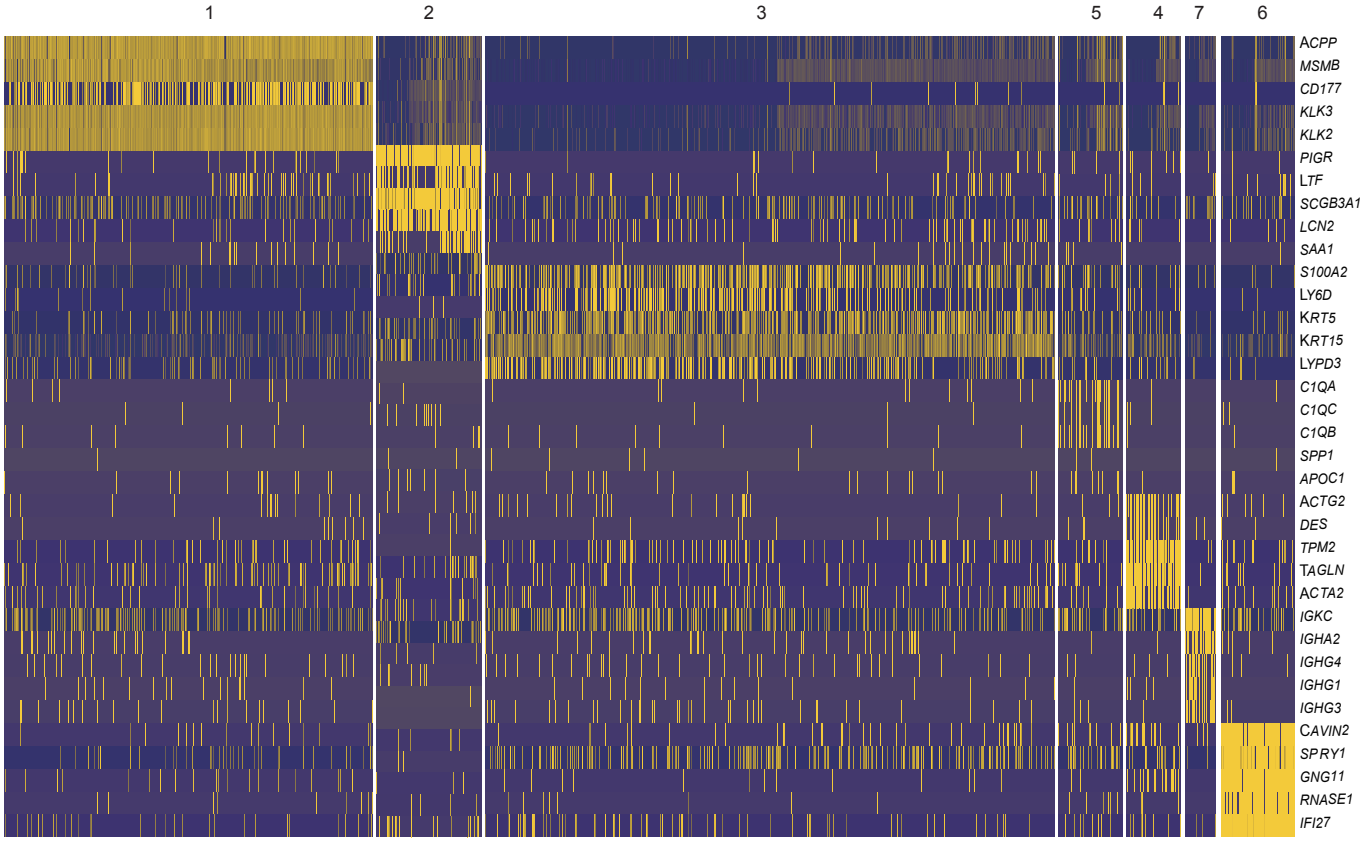
Supplementary Fig. 11 | Location of Luminal-C cells in the regressed C4T mouse prostate. a, Time-line for Luminal-C cell labeling and prostate regression of C4T mice. **b,c,** Co-Immunofluorescence of Ck4, Ck8 with endogenous tdTomato in C4T mouse prostate distal regions (**b**) and proximal regions (**c**). **d,e,** Co-Immunofluorescence of Nkx3.1, Hoxb13 with endogenous tdTomato in C4T mouse prostate distal regions (**d**) and proximal regions (**e**). 6 independent mice were used for each experiment. Scale bars, 50 μm (**b-e**).

Supplementary Figure 12



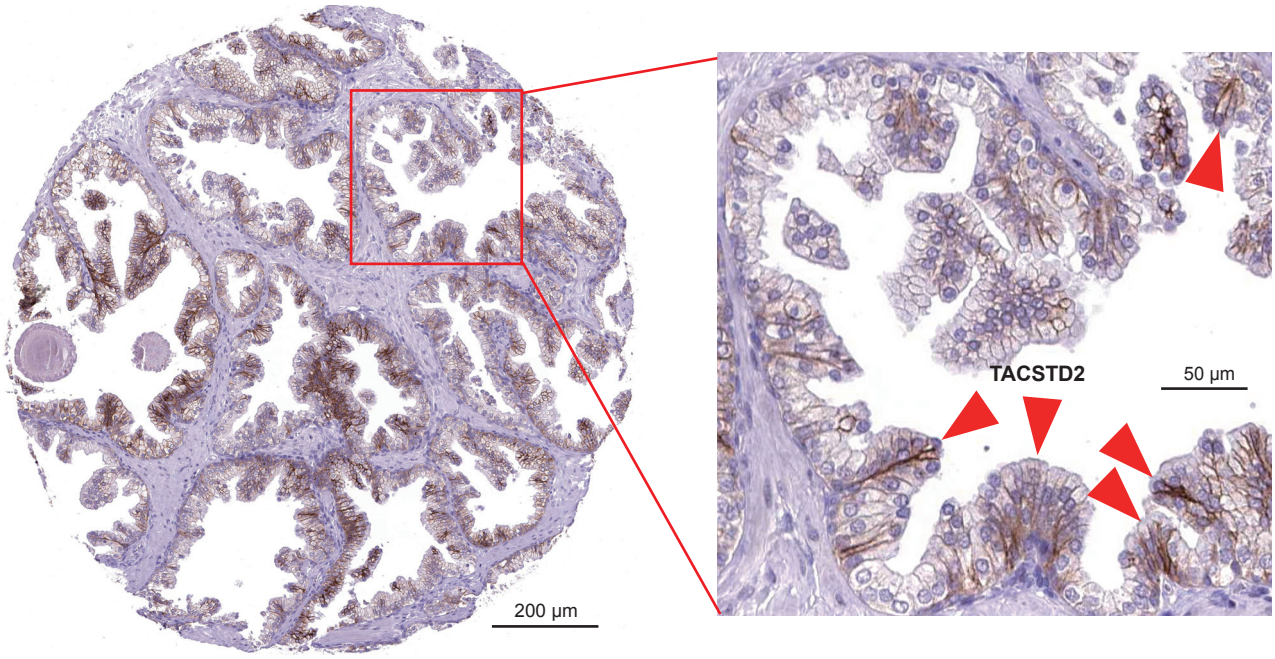
Supplementary Fig. 12 | Human prostate cells distribution of prostate central zone, peripheral zone and transition zone. a, T-SNE map shows the cell distribution from three different human prostate zones (n = 13,744 cells). **b,** Bar graph shows the percentage of the three different prostate zones. 1-7 represents 7 different cell clusters and the colors represent each prostate zone.

Supplementary Figure 13



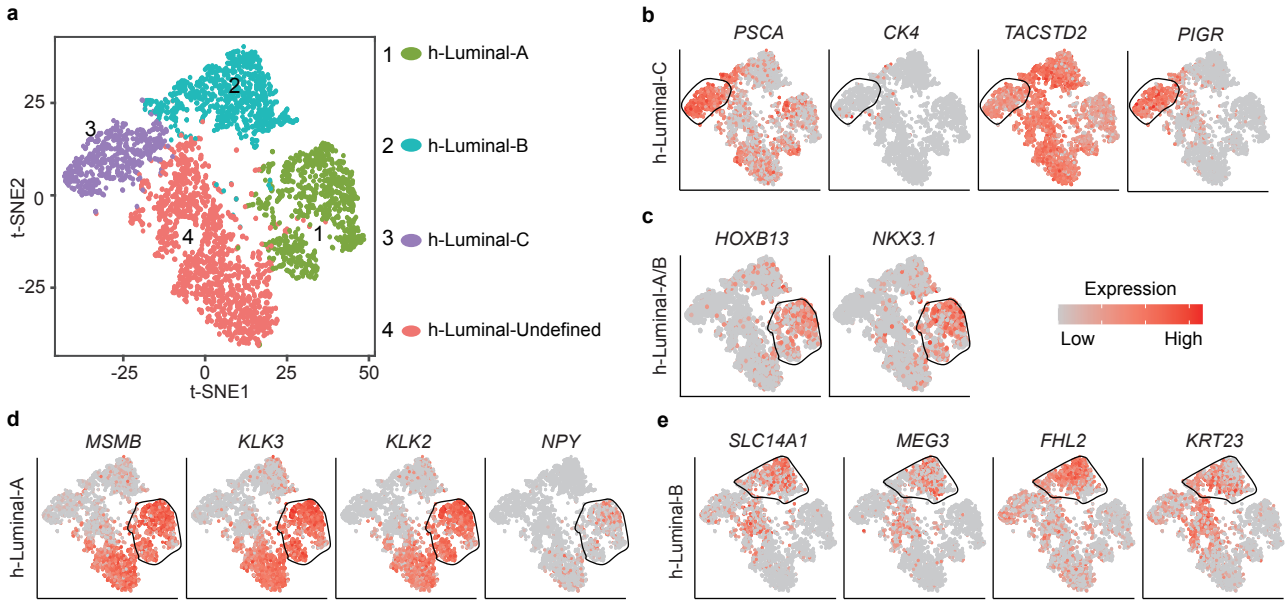
Supplementary Fig. 13 | Specific gene expression patterns for each human prostate cell types. Heatmap shows the relative expression level (row Z-score of log scale normalized read count) of human prostate cells marker genes (rows) of each cluster across cells (columns; n = 11,374 cells).

Supplementary Figure 14



Supplementary Fig. 14 | Location of Luminal-C cells in human normal prostate. TACSTD2-positive luminal cells located at invagination tips of human normal prostate. Scale bars, left 200 μm and right 50 μm.

Supplementary Figure 15



Supplementary Fig. 15 | Decoding of the identity of luminal cell types in the published human prostate atlas. a, Visualization of clustering of prostate single luminal cells, based on the expression of known marker genes by t-SNE (n = 4,039 cells). Decoding of the identity of cell types within the main luminal cell clusters. **b**, T-SNE maps show the expression levels of h-Luminal-C cell markers *CK4*, *PSCA*, *TACSTD2* and *PIGR* (n = 608 cells). **c**, T-SNE maps show the expression levels of mouse Luminal-A/B cell markers *HOXB13* and *NKX3.1* (n = 979 cells). **d**, T-SNE maps show the expression levels of h-Luminal-A cell markers *MSMB*, *KLK3*, *KLK2* and *NPY* (n = 979 cells). **e**, T-SNE maps show the expression levels of Luminal-B cell markers *SLC14A1*, *MEG3*, *FHL2* and *KRT23* (n = 889 cells). Black circle indicate the indicated cell clusters. t-SNE maps showed cells that are colored by the log-scale normalized read count of marker genes.