Figure S1

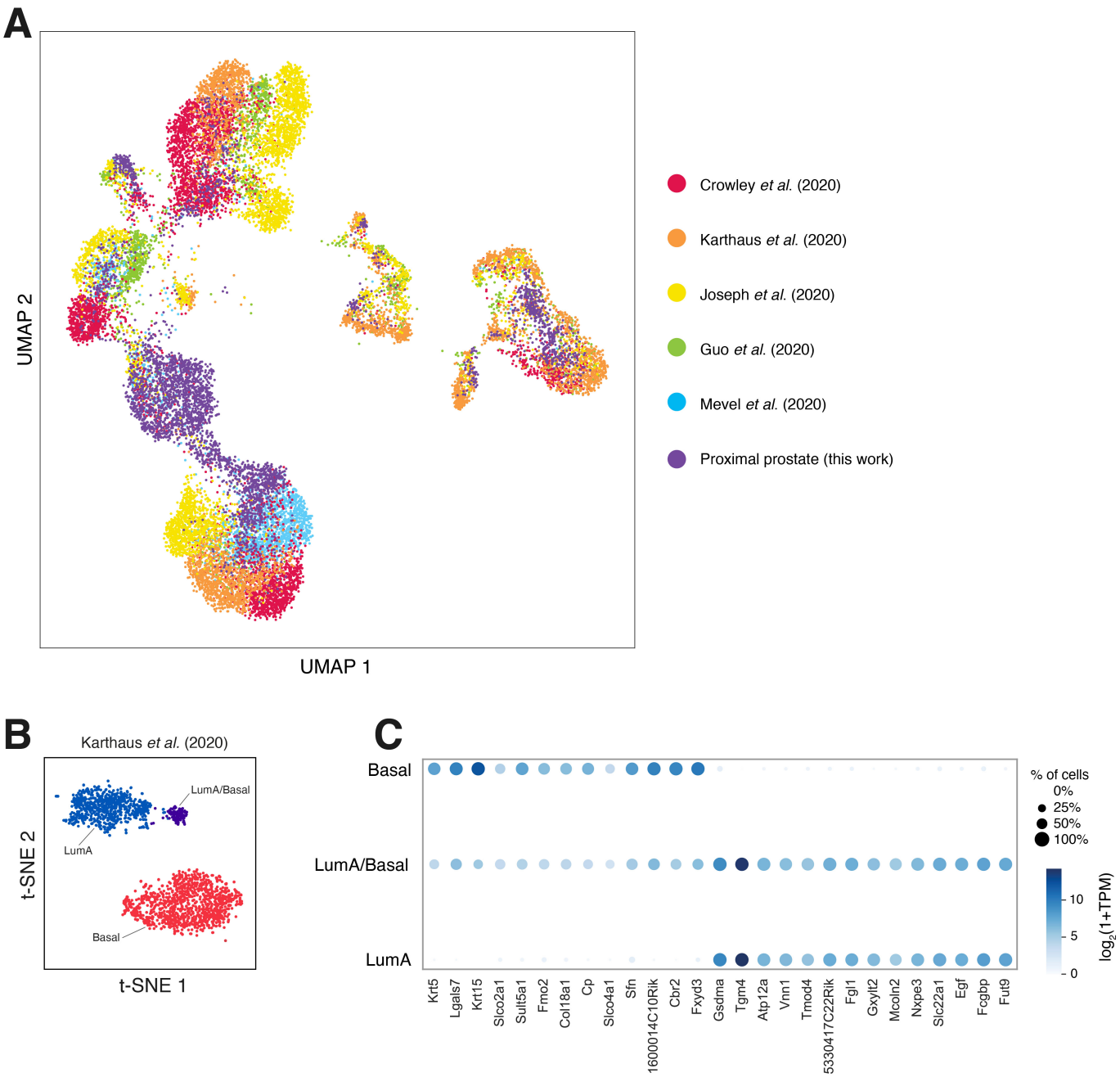

A
UMAP 2
UMAP 1
Crowley et al. (2020)
Karthaus et al. (2020)
Joseph et al. (2020)
Guo et al. (2020)
Mevel et al. (2020)
Proximal prostate (this work)
B
Karthaus et al. (2020)
LumA/Basal
LumA
Basal
t-SNE 2
t-SNE 1
C
Basal
LumA/Basal
LumA
% of cells
0%
25%
50%
100%
log2(1+TPM)
10
5
0
Krt5
Lgals7
Krt15
Slco2a1
Sult5a1
Fmo2
Col18a1
Cp
Slco4a1
Sfn
1600014C10Rik
Cbr2
Fxyd3
Gsdma
Tgm4
Atp12a
Vnn1
Tmod4
5330417C22Rik
Fgl1
Gxylt2
Mcoln2
Nxpe3
Slc22a1
Egf
Fcgbp
Fut9

Fig. S1 Additional analysis of the normal mouse prostate atlas. **A** Alternative visualization of the aggregated dataset in Fig. 1A to show the full contribution of each source dataset to each population. Non-prostatic populations as well as populations that may correspond to cell states are only shown for the individual datasets have been removed. **B,C** Additional subclustering highlights a potential intermediate LumA/Basal cell state in a specific dataset [6], as shown by a UMAP plot (B) and dot plot of detailed marker expression (C).

Figure S2

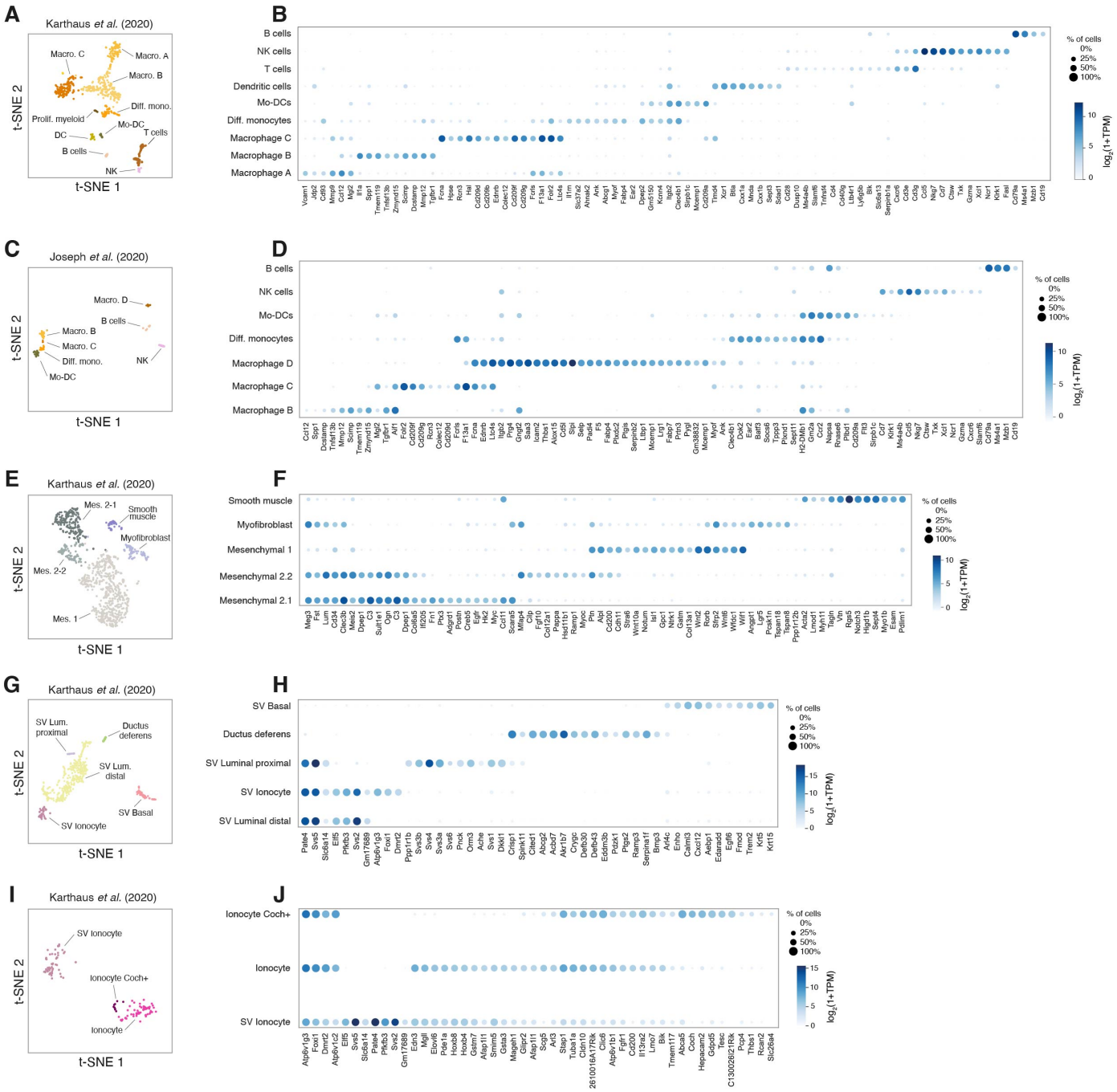
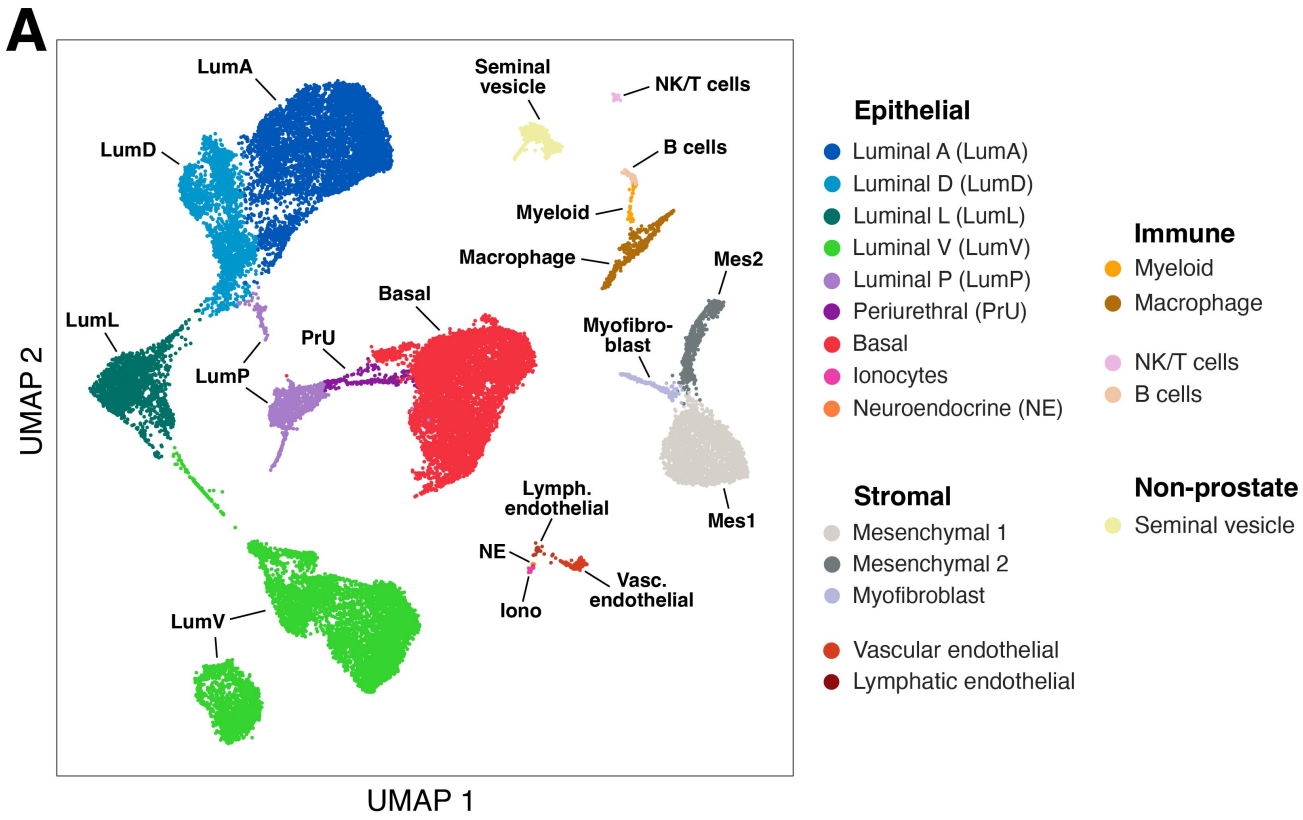
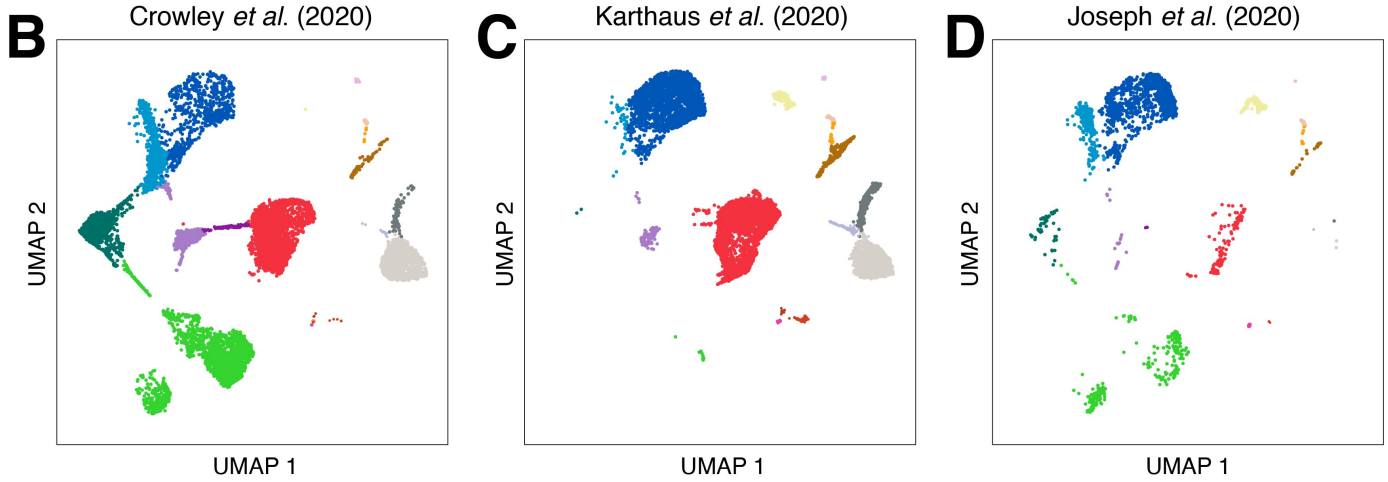

A
Karthaus et al. (2020)
Macro. C
Macro. A
Macro. B
Diff. mono.
Prolif. myeloid
Mo-DC
DC
T cells
B cells
NK
t-SNE 2
t-SNE 1
B
B cells
NK cells
T cells
Dendritic cells
Mo-DCs
Diff. monocytes
Macrophage C
Macrophage B
Macrophage A
% of cells
0%
25%
50%
100%
log2(1+TPM)
C
Joseph et al. (2020)
Macro. D
B cells
Macro. B
Macro. C
Diff. mono.
NK
Mo-DC
t-SNE 2
t-SNE 1
D
B cells
NK cells
Mo-DCs
Diff. monocytes
Macrophage D
Macrophage C
Macrophage B
% of cells
0%
25%
50%
100%
log2(1+TPM)
E
Karthaus et al. (2020)
Mes. 2-1
Smooth muscle
Myofibroblast
Mes. 2-2
Mes. 1
t-SNE 2
t-SNE 1
F
Smooth muscle
Myofibroblast
Mesenchymal 1
Mesenchymal 2.2
Mesenchymal 2.1
% of cells
0%
25%
50%
100%
log2(1+TPM)
G
Karthaus et al. (2020)
SV Lum. proximal
Ductus deferens
SV Lum. distal
SV Basal
SV Ionocyte
t-SNE 2
t-SNE 1
H
SV Basal
Ductus deferens
SV Luminal proximal
SV Ionocyte
SV Luminal distal
% of cells
0%
25%
50%
100%
log2(1+TPM)
I
Karthaus et al. (2020)
SV Ionocyte
Ionocyte Coch+
Ionocyte
t-SNE 2
t-SNE 1
J
Ionocyte Coch+
Ionocyte
SV Ionocyte
% of cells
0%
25%
50%
100%
log2(1+TPM)

Fig. S2 Additional subclustering of specific cell populations from the prostate and neighboring tissues. **A,B** The immune compartment from [6] contains at least 5 distinct myeloid populations (A); dot plot shows distinguishing markers (B). **C,D** The immune compartment from [7] contains at least 4 distinct myeloid populations (C); dot plot showing distinguishing markers (D). **E,F** The stromal compartment from [6] contains 4 distinct populations with potential further subdivision of the Mes2 population. **G,H** The non-prostate cells from [6] primarily capture seminal vesicle and reveal potential proximal/distal subdivision of the tissue. **I,J** The Foxi1-positive ionocytes from [6] show potential heterogeneity as well and are present in both prostate and seminal vesicle tissue.


A
LumA
LumD
LumL
LumP
PrU
Basal
LumV
Seminal vesicle
NK/T cells
B cells
Myeloid
Macrophage
Mes2
Myofibro-blast
Mes1
Lymph. endothelial
NE
Iono
Vasc. endothelial
UMAP 2
UMAP 1
Epithelial
Luminal A (LumA)
Luminal D (LumD)
Luminal L (LumL)
Luminal V (LumV)
Luminal P (LumP)
Periurethral (PrU)
Basal
Ionocytes
Neuroendocrine (NE)
Stromal
Mesenchymal 1
Mesenchymal 2
Myofibroblast
Vascular endothelial
Lymphatic endothelial
Immune
Myeloid
Macrophage
NK/T cells
B cells
Non-prostate
Seminal vesicle


B
Crowley et al. (2020)
C
Karthaus et al. (2020)
D
Joseph et al. (2020)
UMAP 2
UMAP 1

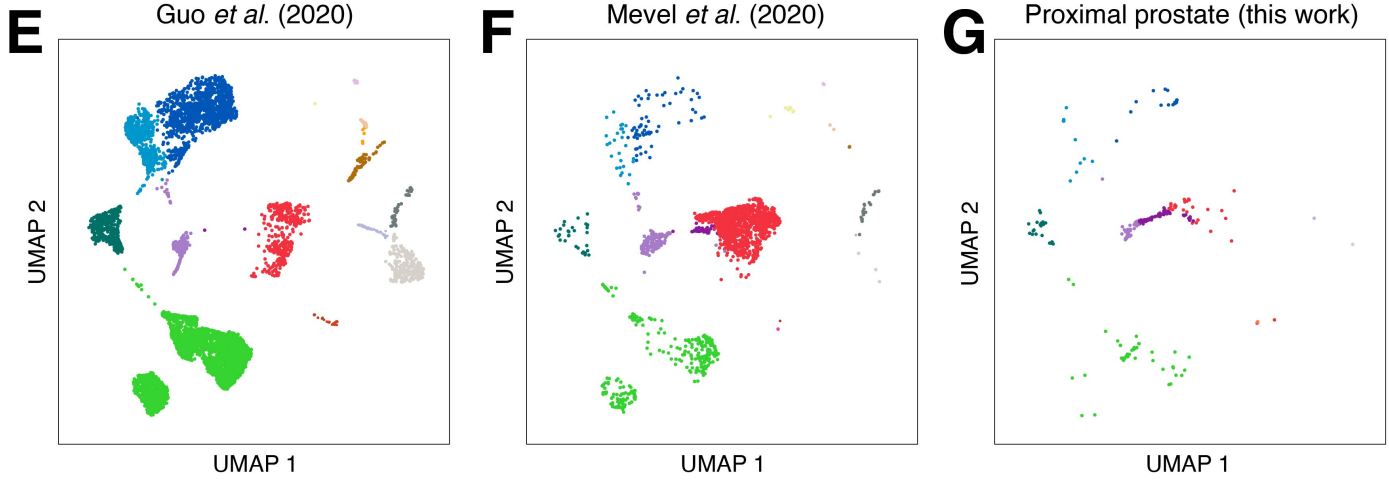

E
Guo et al. (2020)
F
Mevel et al. (2020)
G
Proximal prostate (this work)
UMAP 2
UMAP 1

Fig. S3 Reference plots of mouse prostate scRNA-seq data generated by the Seurat pipeline. **A** Shown are plots arranged as in Fig. 1, revealing 9 epithelial populations, 5 stromal populations, and at least 4 immune populations across multiple datasets. This approach reproduced all epithelial and stromal populations; however, it did not subcluster some low abundance and immune populations, so these were labeled as combined populations. **B-G** Sources of the datasets that compose the aggregate UMAP, showing intermixing of cells from different mouse strains and processing methods within population clusters.

Figure S4

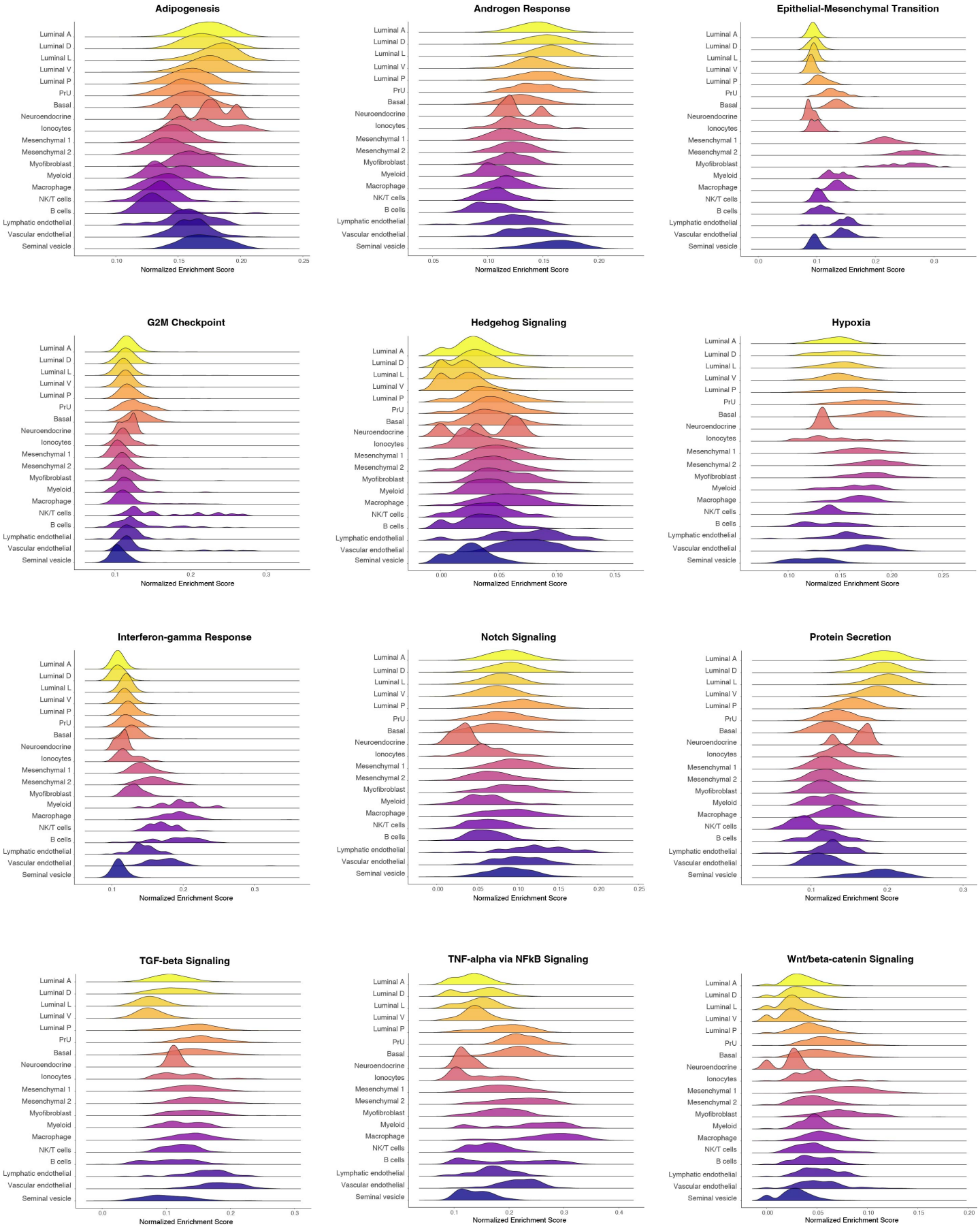

Adipogenesis
Androgen Response
Epithelial-Mesenchymal Transition
G2M Checkpoint
Hedgehog Signaling
Hypoxia
Interferon-gamma Response
Notch Signaling
Protein Secretion
TGF-beta Signaling
TNF-alpha via NFkB Signaling
Wnt/beta-catenin Signaling
Luminal A
Luminal D
Luminal L
Luminal V
Luminal P
PrU
Basal
Neuroendocrine
Ionocytes
Mesenchymal 1
Mesenchymal 2
Myofibroblast
Myeloid
Macrophage
NK/T cells
B cells
Lymphatic endothelial
Vascular endothelial
Seminal vesicle
Normalized Enrichment Score

Fig. S4 Gene Set Enrichment Analysis of the mouse prostate populations using Hallmark Pathways. Pathway analysis of curated gene expression signatures shows the normalized expression of gene sets involved in different processes by cell type. Of note, among the prostate epithelial populations, distal luminal cells are the most enriched for genes involved in protein secretion and androgen response, whereas proximal cell types display more enrichment for Wnt, TGF-beta, and TNF-alpha pathway genes. Analyses for each pathway were performed on aggregated mouse prostate data in the Seurat pipeline.

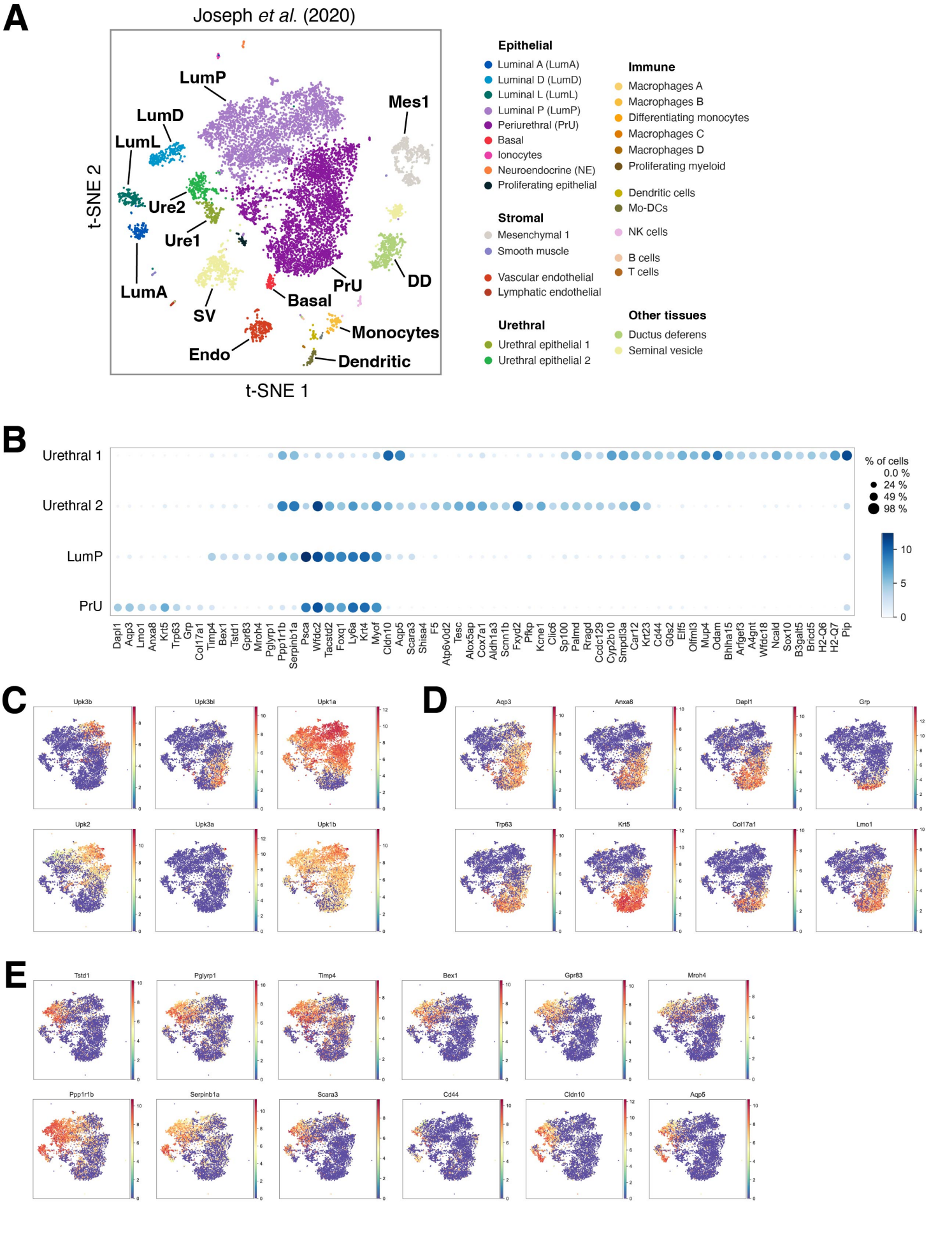

A
Joseph et al. (2020)
LumP
LumD
LumL
Ure2
Ure1
LumA
SV
Endo
Basal
PrU
DD
Mes1
Monocytes
Dendritic
t-SNE 2
t-SNE 1
Epithelial
Luminal A (LumA)
Luminal D (LumD)
Luminal L (LumL)
Luminal P (LumP)
Periurethral (PrU)
Basal
Ionocytes
Neuroendocrine (NE)
Proliferating epithelial
Stromal
Mesenchymal 1
Smooth muscle
Vascular endothelial
Lymphatic endothelial
Urethral
Urethral epithelial 1
Urethral epithelial 2
Immune
Macrophages A
Macrophages B
Differentiating monocytes
Macrophages C
Macrophages D
Proliferating myeloid
Dendritic cells
Mo-DCs
NK cells
B cells
T cells
Other tissues
Ductus deferens
Seminal vesicle
B
Urethral 1
Urethral 2
LumP
PrU
% of cells
0.0 %
24 %
49 %
98 %
10
5
0
Dapl1 Aqp3 Lmo1 Anxa8 Krt5 Trp63 Grp Col17a1 Timp4 Bex1 Tstd1 Gpr83 Mroh4 Pglyrp1 Ppp1r1b Serpinb1a Psca Wfdc2 Tacstd2 Foxq1 Ly6a Krt4 Myof Cldn10 Aqp5 Scara3 Shisa4 F5 Atp6v0d2 Tesc Alox5ap Cox7a1 Aldh1a3 Scnn1b Fxyd2 Pfkp Kcne1 Clic6 Sp100 Palmd Rragd Ccdc129 Cyp2b10 Smpdl3a Car12 Krt23 Cd44 G0s2 Elf5 Olfml3 Mup4 Odam Bhlha15 Arfgef3 A4gnt Wfdc18 Ncald Sox10 B3galt5 Bricd5 H2-Q6 H2-Q7 Pip
C
Upk3b
Upk3bl
Upk1a
Upk2
Upk3a
Upk1b
D
Aqp3
Anxa8
Dapl1
Grp
Trp63
Krt5
Col17a1
Lmo1
E
Tstd1
Pglyrp1
Timp4
Bex1
Gpr83
Mroh4
Ppp1r1b
Serpinb1a
Scara3
Cd44
Cldn10
Aqp5

Fig. S5 Analysis of a published dataset for the mouse urethral and periurethral regions. **A** t-SNE visualization of the urethral dataset from [7] shows distal, proximal, periurethral, basal, neuroendocrine, and ionocyte cells, mesenchymal and immune populations, and several non-prostate populations such as the urethra. **B** Dot plot of differentially expressed markers between LumP and PrU (mixed), and urethra-only populations in the urethral dataset. Although they are visibly distinct, both the prostate PrU and LumP populations have gene expression overlap with similar urethral populations. **C-E** Genes with enriched expression in the urothelium (C), prostatic PrU (D), and prostatic LumP (E).

Figure S6

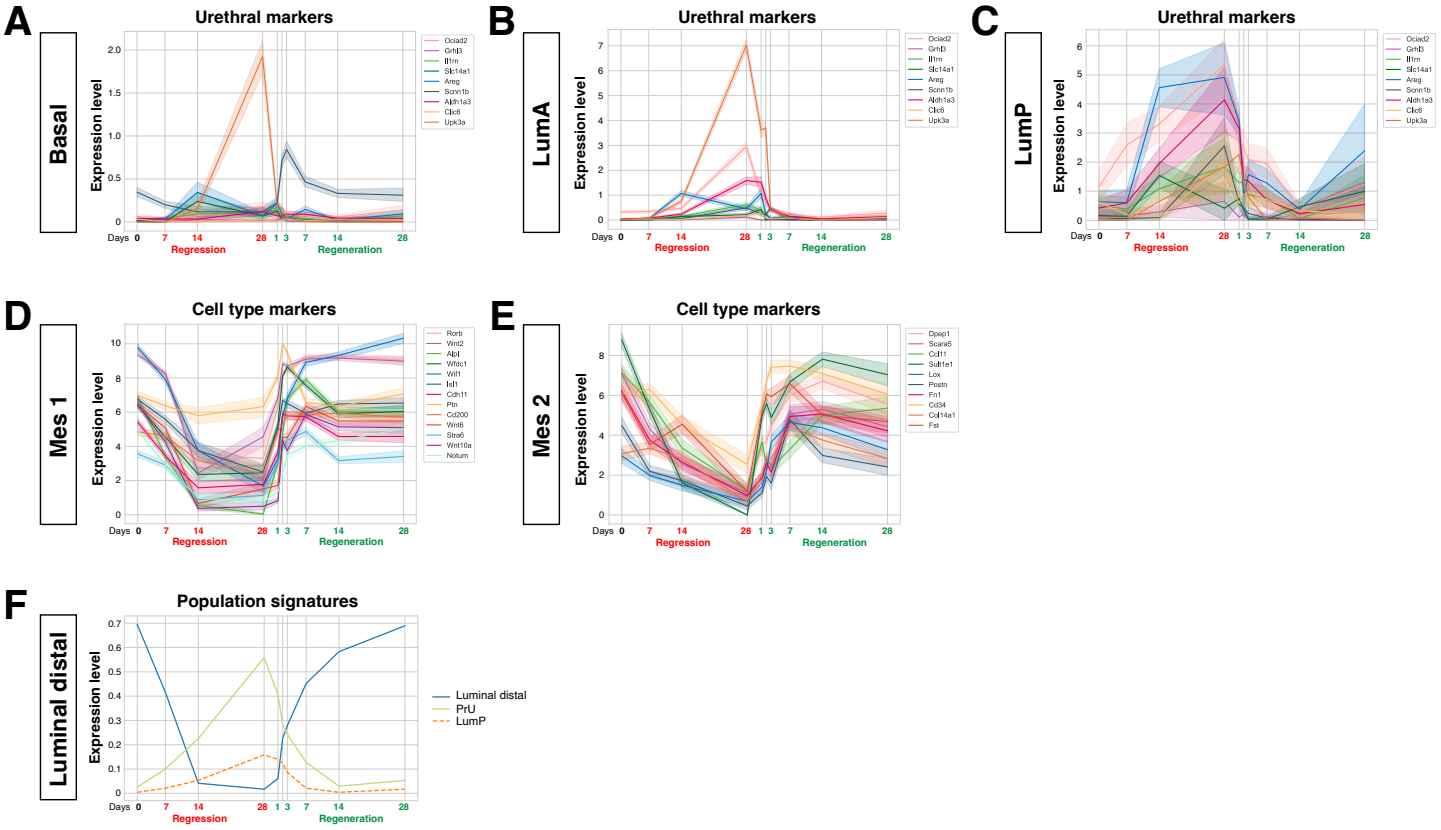

A
Basal
Urethral markers
B
LumA
Urethral markers
C
LumP
Urethral markers
Expression level
Days 0 7 14 28 1 3 7 14 28
Regression
Regeneration
Ociad2
Grhl3
Ifitm
Slc14a1
Areg
Scnn1b
Aldh1a3
Clic6
Upk3a
D
Mes 1
Cell type markers
E
Mes 2
Cell type markers
Rorb
Wnt2
Alpl
Wfdc1
Wif1
Isl1
Cdh11
Ptn
Cd200
Wnt6
Strc6
Wnt10a
Notum
Dpep1
Scara5
Cd11
Sult1e1
Lox
Postn
Fn1
Cd34
Cd14a1
Fst
F
Luminal distal
Population signatures
Luminal distal
PrU
LumP

Fig. S6 Gene expression changes during prostate regression and regeneration. Line plots generated from meta-analysis of a regression-regeneration time course based on [6], where the line indicates average gene expression across the population and the bar indicates the CI (+/- 95%). **A-C** Genes that are enriched in urethral and not PrU cells, in Basal (A), LumA (B), and LumP (C) cells. LumA and Basal cells generally did not gain expression of urothelial markers, with the exception of Upk3a, which is specific to the urethra in C57BL/6 mice, whereas LumP cells gained expression of a subset of urothelial markers in addition to the PrU markers. **D,E** Expression of distinguishing markers for the Mes 1 (D) and Mes 2 (E) fibroblast populations over the regression and regeneration time course.

Figure S7

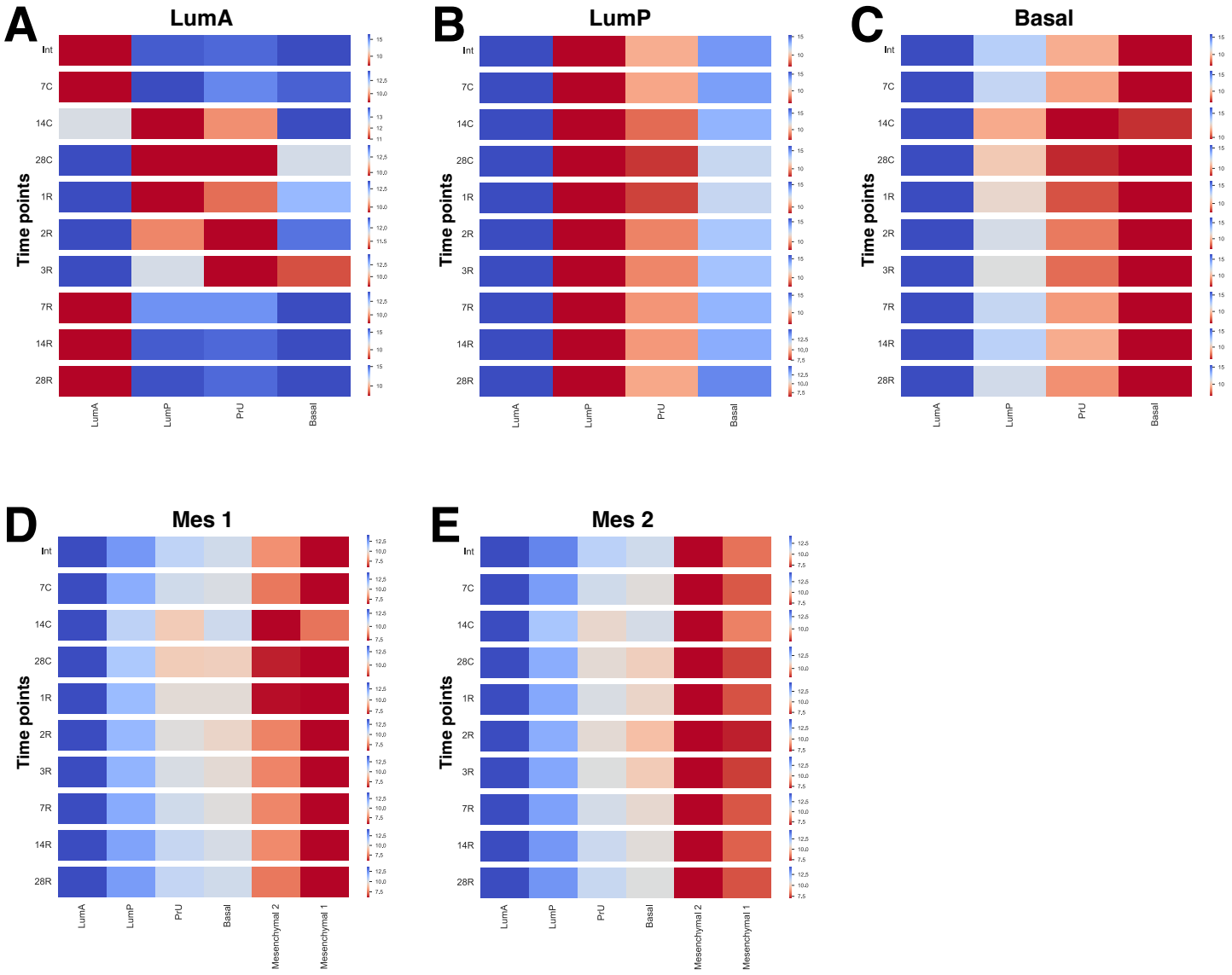

A
LumA
B
LumP
C
Basal
D
Mes 1
E
Mes 2
Time points
Int
7C
14C
28C
1R
2R
3R
7R
14R
28R
LumA
LumP
PrU
Basal
Mesenchymal 2
Mesenchymal 1

Fig. S7 Heatmaps of total gene expression profiles for individual cell populations during prostate regression and regeneration. Profiles of LumA (A), LumP (B), Basal (C), as well as Mes 1 (D) and Mes 2 (E) based on data from [6] are compared to the normal gene expression profiles from the Tabula Sapiens. Measurements are given in Wasserstein distance, a metric of effort required to convert one profile into the other; red indicates greater similarity, and blue indicates less similarity between populations.

Figure S8

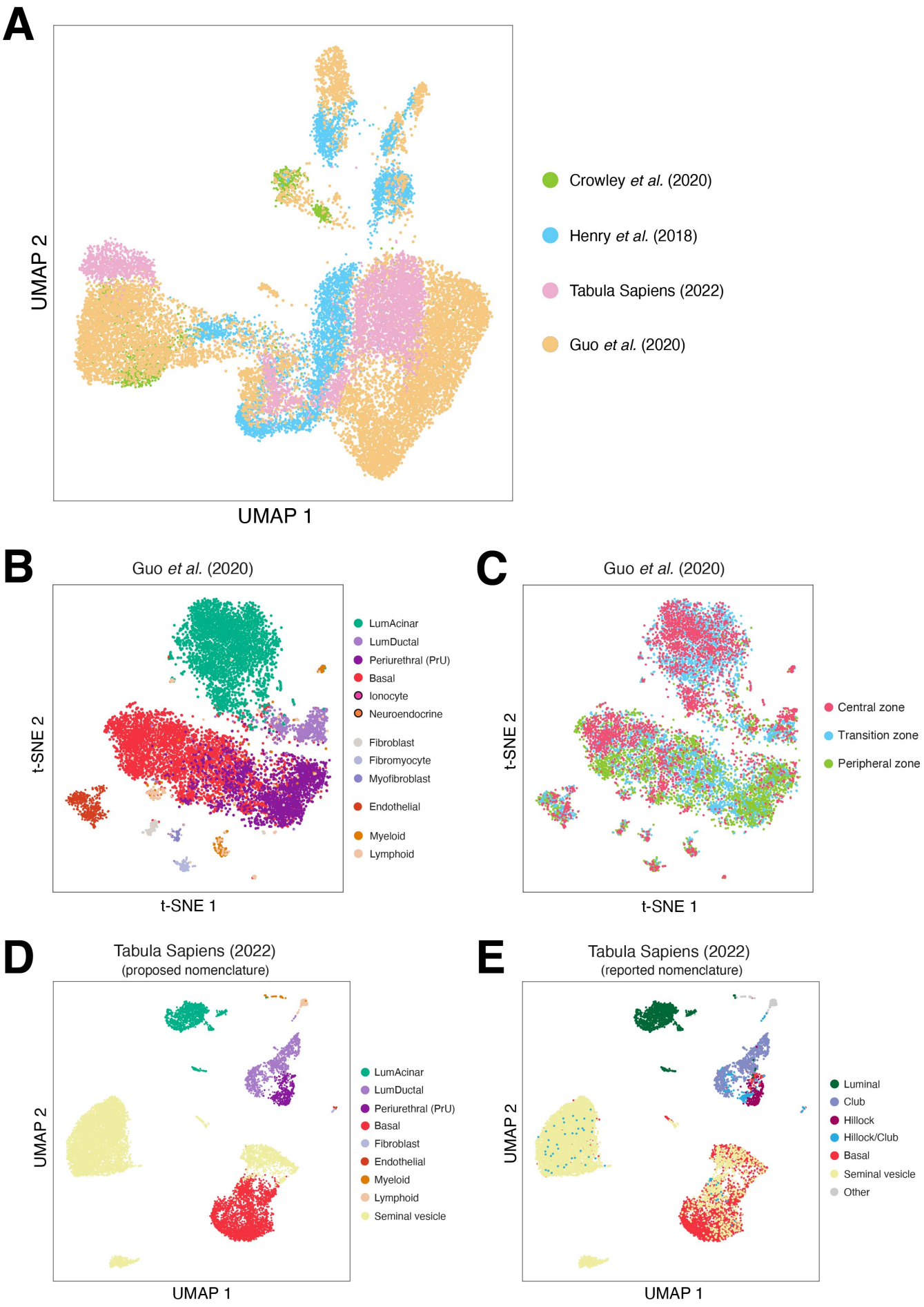

A
UMAP 2
UMAP 1
Crowley et al. (2020)
Henry et al. (2018)
Tabula Sapiens (2022)
Guo et al. (2020)
B
Guo et al. (2020)
t-SNE 2
t-SNE 1
LumAcinar
LumDuctal
Periurethral (PrU)
Basal
Ionocyte
Neuroendocrine
Fibroblast
Fibromyocyte
Myofibroblast
Endothelial
Myeloid
Lymphoid
C
Guo et al. (2020)
t-SNE 2
t-SNE 1
Central zone
Transition zone
Peripheral zone
D
Tabula Sapiens (2022)
(proposed nomenclature)
UMAP 2
UMAP 1
LumAcinar
LumDuctal
Periurethral (PrU)
Basal
Fibroblast
Endothelial
Myeloid
Lymphoid
Seminal vesicle
E
Tabula Sapiens (2022)
(reported nomenclature)
UMAP 2
UMAP 1
Luminal
Club
Hillock
Hillock/Club
Basal
Seminal vesicle
Other

Fig. S8 Additional analysis of the normal human prostate atlas. **A** Alternative visualization of the aggregated dataset in Fig. 4A showing the source datasets. **B,C** UMAP of the data from [5] using our descriptive nomenclature as shown in Fig. 4C (B), and the zonal identity of each cell (C). Although each large epithelial population contains cells from each zone, proportionally fewer of the LumAcinar cells were from the peripheral zone. **D, E** Human prostate nomenclature shown for the Tabula Sapiens normal prostate dataset [13], comparing the proposed nomenclature (D) to the nomenclature from [11] (E).

Figure S9

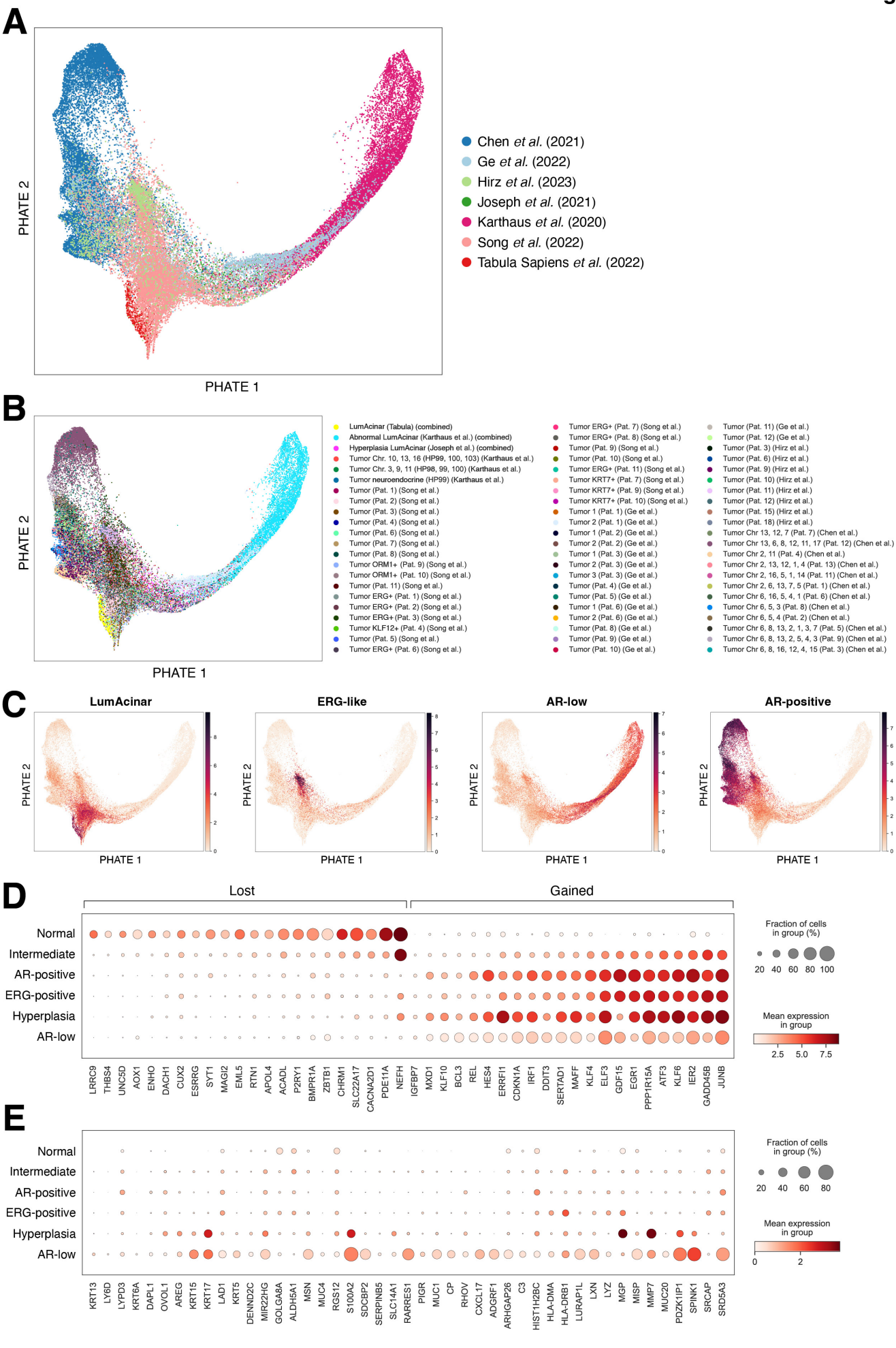

A
PHATE 2
PHATE 1
Chen et al. (2021)
Ge et al. (2022)
Hirz et al. (2023)
Joseph et al. (2021)
Karthaus et al. (2020)
Song et al. (2022)
Tabula Sapiens et al. (2022)
B
PHATE 2
PHATE 1
LumAcinar (Tabula) (combined)
Abnormal LumAcinar (Karthaus et al.) (combined)
Hyperplasia LumAcinar (Joseph et al.) (combined)
Tumor Chr. 10, 13, 16 (HP99, 100, 103) (Karthaus et al.)
Tumor Chr. 3, 9, 11 (HP98, 99, 100) (Karthaus et al.)
Tumor neuroendocrine (HP99) (Karthaus et al.)
Tumor (Pat. 1) (Song et al.)
Tumor (Pat. 2) (Song et al.)
Tumor (Pat. 3) (Song et al.)
Tumor (Pat. 4) (Song et al.)
Tumor (Pat. 6) (Song et al.)
Tumor (Pat. 7) (Song et al.)
Tumor (Pat. 8) (Song et al.)
Tumor ORM1+ (Pat. 9) (Song et al.)
Tumor ORM1+ (Pat. 10) (Song et al.)
Tumor (Pat. 11) (Song et al.)
Tumor ERG+ (Pat. 1) (Song et al.)
Tumor ERG+ (Pat. 2) (Song et al.)
Tumor ERG+ (Pat. 3) (Song et al.)
Tumor KLF12+ (Pat. 4) (Song et al.)
Tumor (Pat. 5) (Song et al.)
Tumor ERG+ (Pat. 6) (Song et al.)
Tumor ERG+ (Pat. 7) (Song et al.)
Tumor ERG+ (Pat. 8) (Song et al.)
Tumor (Pat. 9) (Song et al.)
Tumor (Pat. 10) (Song et al.)
Tumor ERG+ (Pat. 11) (Song et al.)
Tumor KRT7+ (Pat. 7) (Song et al.)
Tumor KRT7+ (Pat. 9) (Song et al.)
Tumor KRT7+ (Pat. 10) (Song et al.)
Tumor 1 (Pat. 1) (Ge et al.)
Tumor 2 (Pat. 1) (Ge et al.)
Tumor 1 (Pat. 2) (Ge et al.)
Tumor 2 (Pat. 2) (Ge et al.)
Tumor 1 (Pat. 3) (Ge et al.)
Tumor 2 (Pat. 3) (Ge et al.)
Tumor 3 (Pat. 3) (Ge et al.)
Tumor (Pat. 4) (Ge et al.)
Tumor (Pat. 5) (Ge et al.)
Tumor 1 (Pat. 6) (Ge et al.)
Tumor 2 (Pat. 6) (Ge et al.)
Tumor (Pat. 8) (Ge et al.)
Tumor (Pat. 9) (Ge et al.)
Tumor (Pat. 10) (Ge et al.)
Tumor (Pat. 11) (Ge et al.)
Tumor (Pat. 12) (Ge et al.)
Tumor (Pat. 3) (Hirz et al.)
Tumor (Pat. 6) (Hirz et al.)
Tumor (Pat. 9) (Hirz et al.)
Tumor (Pat. 10) (Hirz et al.)
Tumor (Pat. 11) (Hirz et al.)
Tumor (Pat. 12) (Hirz et al.)
Tumor (Pat. 15) (Hirz et al.)
Tumor (Pat. 18) (Hirz et al.)
Tumor Chr 13, 12, 7 (Pat. 7) (Chen et al.)
Tumor Chr 13, 6, 8, 12, 11, 17 (Pat. 12) (Chen et al.)
Tumor Chr 2, 11 (Pat. 4) (Chen et al.)
Tumor Chr 2, 13, 12, 1, 4 (Pat. 13) (Chen et al.)
Tumor Chr 2, 16, 5, 1, 14 (Pat. 11) (Chen et al.)
Tumor Chr 2, 6, 13, 7, 5 (Pat. 1) (Chen et al.)
Tumor Chr 6, 16, 5, 4, 1 (Pat. 6) (Chen et al.)
Tumor Chr 6, 5, 3 (Pat. 8) (Chen et al.)
Tumor Chr 6, 5, 4 (Pat. 2) (Chen et al.)
Tumor Chr 6, 8, 13, 2, 1, 3, 7 (Pat. 5) (Chen et al.)
Tumor Chr 6, 8, 13, 2, 5, 4, 3 (Pat. 9) (Chen et al.)
Tumor Chr 6, 8, 16, 12, 4, 15 (Pat. 3) (Chen et al.)
C
LumAcinar
ERG-like
AR-low
AR-positive
PHATE 2
PHATE 1
D
Lost
Gained
Normal
Intermediate
AR-positive
ERG-positive
Hyperplasia
AR-low
LRRC9 THBS4 UNC5D AOX1 ENHO DACH1 CUX2 ESRRG SYT1 MAGI2 EML5 RTN1 APOL4 ACADL P2RY1 BMPR1A ZBTB1 CHRM1 SLC22A17 CACNA2D1 PDE11A NEFH IGFBP7 MXD1 KLF10 BCL3 REL HES4 ERRFI1 CDKN1A IRF1 DDIT3 SERTAD1 MAFF KLF4 ELF3 GDF15 EGR1 PPP1R15A ATF3 KLF6 IER2 GADD45B JUNB
Fraction of cells in group (%)
20 40 60 80 100
Mean expression in group
2.5 5.0 7.5
E
Normal
Intermediate
AR-positive
ERG-positive
Hyperplasia
AR-low
KRT13 LY6D LYPD3 KRT6A DAPL1 OVOL1 AREG KRT15 KRT17 LAD1 KRT5 DENND2C MIR22HG GOLGA8A ALDH5A1 MSN MUC4 RGS12 S100A2 SDCBP2 SERPINB5 SLC14A1 RARRES1 PIGR MUC1 CP RHOV CXCL17 ADGRF1 ARHGAP26 C3 HIST1H2BC HLA-DMA HLA-DRB1 LURAP1L LXN LYZ MGP MISP MMP7 MUC20 PDZK1IP1 SPINK1 SRCAP SRD5A3
Fraction of cells in group (%)
20 40 60 80
Mean expression in group
0 2

Fig. S9 Gene expression profiles and cohort details for LumAcinar cells from normal and transformed prostates. **A-C** PHATE plots of LumAcinar cells corresponding to Fig. 5A. **A** Data contribution from the 6 cohorts [6, 7, 13-17]. **B** Disease state, subgroup, and tumor samples for the patients within each cohort. CNV analyses are detailed in the Methods and in Fig. 6. **C** Average expression of the gene sets associated with normal and transformed LumAcinar cell states. Transformed and abnormal cells exhibit heterogeneous profiles that divide into those with high androgen receptor signaling (AR-positive), ERG-positive, or low AR signaling. **D** Dot plot displaying selected genes with differential expression between normal and transformed LumAcinar cells, divided into those lost (left) or gained (right) during transformation. **E** Dot plot of selected genes from the PrU and LumDuctal signatures.

Figure S10

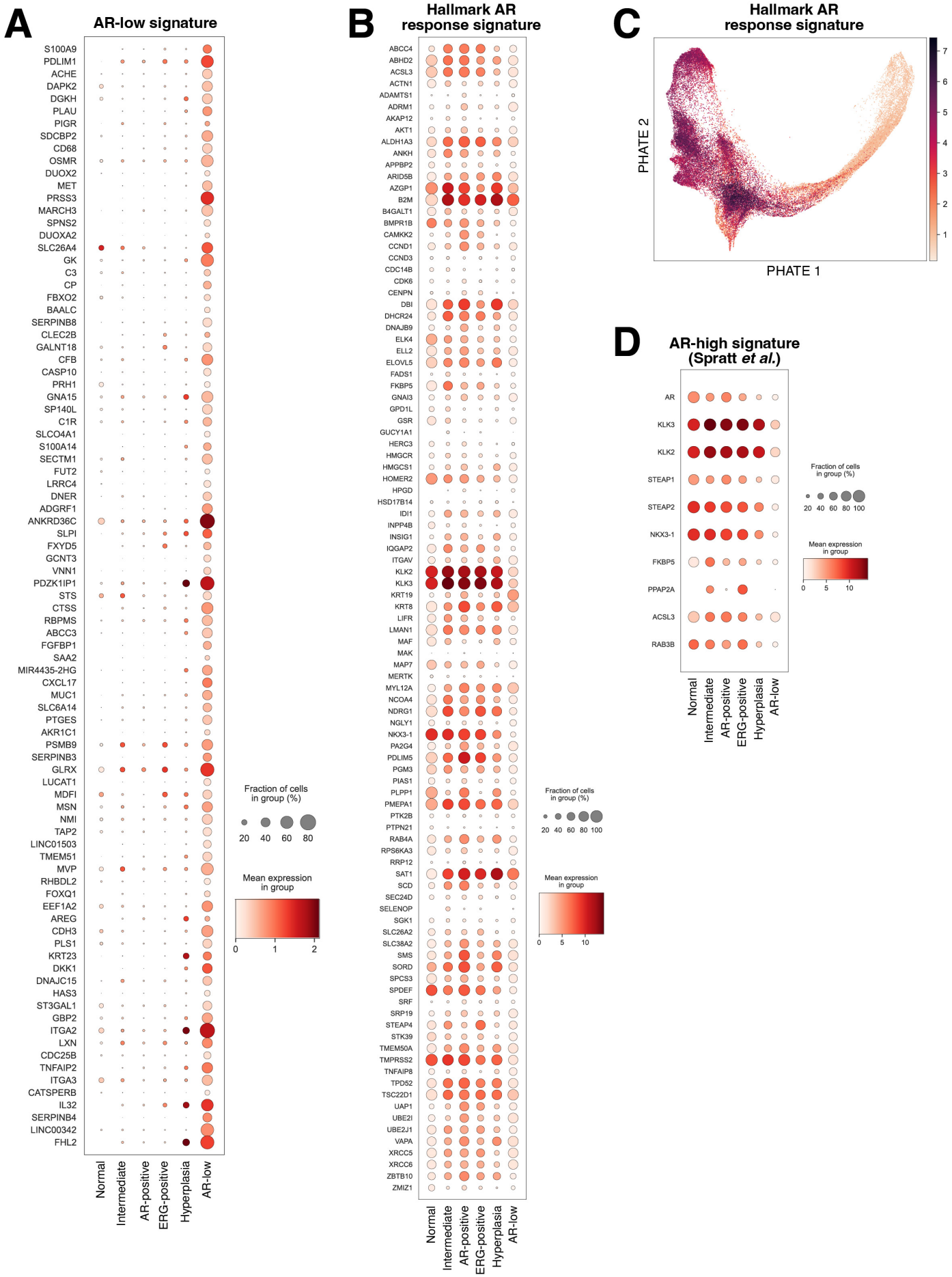
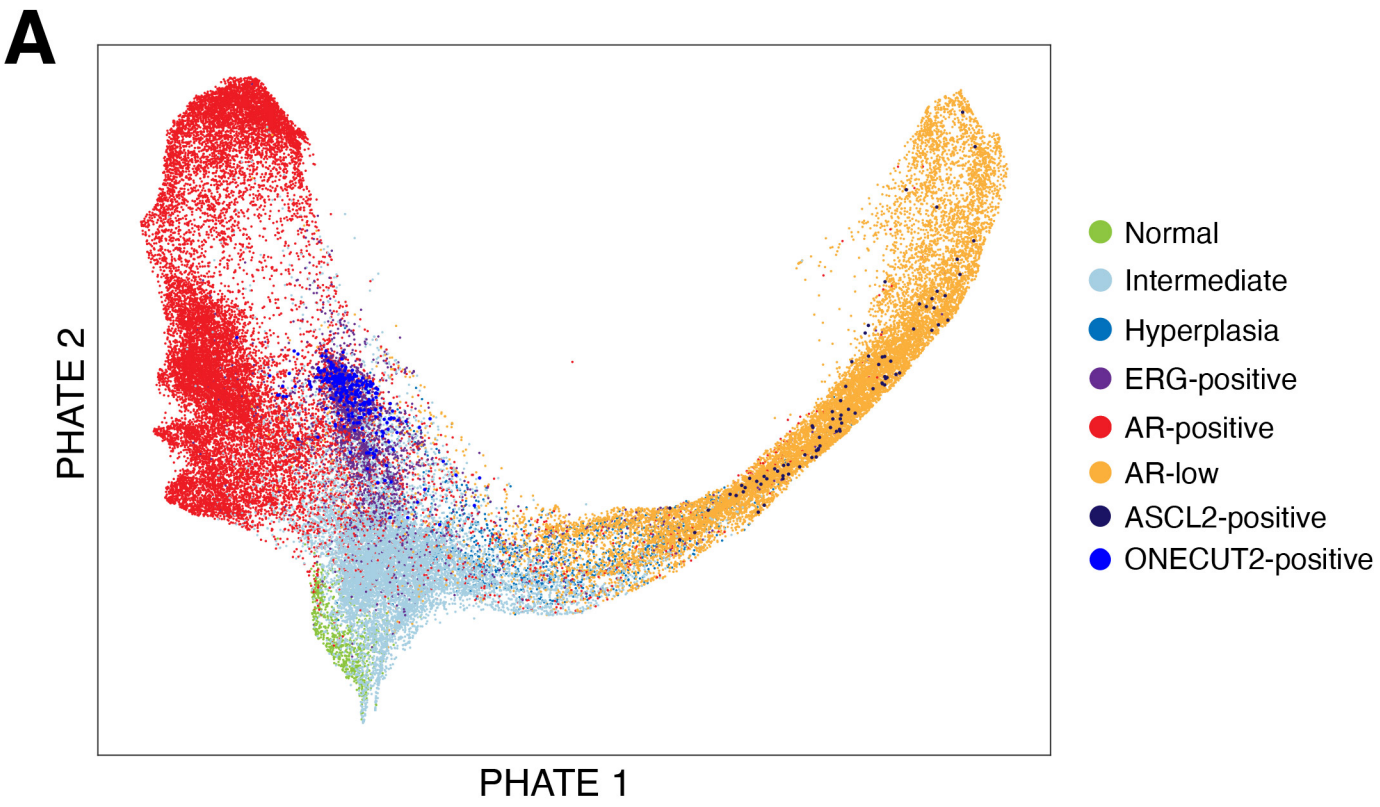
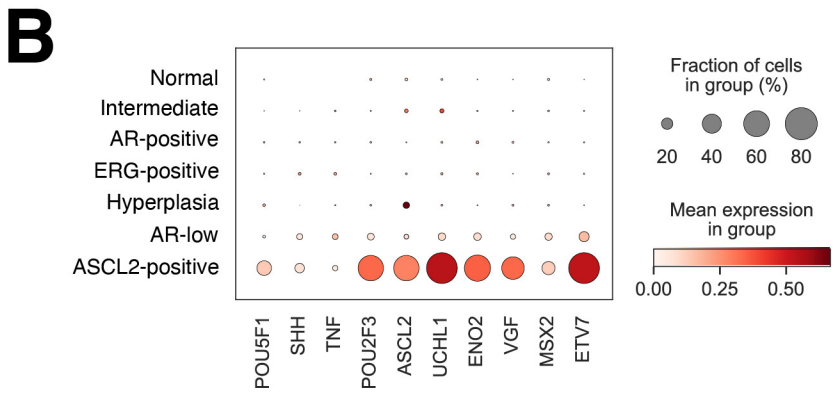

A
AR-low signature
S100A9
PDLIM1
ACHE
DAPK2
DGKH
PLAU
PIGR
SDCBP2
CD68
OSMR
DUOX2
MET
PRSS3
MARCH3
SPNS2
DUOXA2
SLC26A4
GK
C3
CP
FBXO2
BAALC
SERPINB8
CLEC2B
GALNT18
CFB
CASP10
PRH1
GNA15
SP140L
C1R
SLCO4A1
S100A14
SECTM1
FUT2
LRRC4
DNER
ADGRF1
ANKRD36C
SLPI
FXYD5
GCNT3
VNN1
PDZK1IP1
STS
CTSS
RBPMS
ABCC3
FGFBP1
SAA2
MIR4435-2HG
CXCL17
MUC1
SLC6A14
PTGES
AKR1C1
PSMB9
SERPINB3
GLRX
LUCAT1
MDFI
MSN
NMI
TAP2
LINC01503
TMEM51
MVP
RHBDL2
FOXQ1
EEF1A2
AREG
CDH3
PLS1
KRT23
DKK1
DNAJC15
HAS3
ST3GAL1
GBP2
ITGA2
LXN
CDC25B
TNFAIP2
ITGA3
CATSPERB
IL32
SERPINB4
LINC00342
FHL2
Normal
Intermediate
AR-positive
ERG-positive
Hyperplasia
AR-low
Fraction of cells in group (%)
20 40 60 80
Mean expression in group
0 1 2
B
Hallmark AR response signature
ABCC4
ABHD2
ACSL3
ACTN1
ADAMTS1
ADRM1
AKAP12
AKT1
ALDH1A3
ANKH
APPBP2
ARID5B
AZGP1
B2M
B4GALT1
BMPR1B
CAMKK2
CCND1
CCND3
CDC14B
CDK6
CENPN
DBI
DHCR24
DNAJB9
ELK4
ELL2
ELOVL5
FADS1
FKBP5
GNAI3
GPD1L
GSR
GUCY1A1
HERC3
HMGCR
HMGCS1
HOMER2
HPGD
HSD17B14
IDI1
INPP4B
INSIG1
IQGAP2
ITGAV
KLK2
KLK3
KRT19
KRT8
LIFR
LMAN1
MAF
MAK
MAP7
MERTK
MYL12A
NCOA4
NDRG1
NGLY1
NKX3-1
PA2G4
PDLIM5
PGM3
PIAS1
PLPP1
PMEPA1
PTK2B
PTPN21
RAB4A
RPS6KA3
RRP12
SAT1
SCD
SEC24D
SELENOP
SGK1
SLC26A2
SLC38A2
SMS
SORD
SPCS3
SPDEF
SRF
SRP19
STEAP4
STK39
TMEM50A
TMPRSS2
TNFAIP8
TPD52
TSC22D1
UAP1
UBE2I
UBE2J1
VAPA
XRCC5
XRCC6
ZBTB10
ZMIZ1
Normal
Intermediate
AR-positive
ERG-positive
Hyperplasia
AR-low
Fraction of cells in group (%)
20 40 60 80 100
Mean expression in group
0 5 10
C
Hallmark AR response signature
PHATE 2
PHATE 1
7
6
5
4
3
2
1
D
AR-high signature (Spratt et al.)
AR
KLK3
KLK2
STEAP1
STEAP2
NKX3-1
FKBP5
PPAP2A
ACSL3
RAB3B
Normal
Intermediate
AR-positive
ERG-positive
Hyperplasia
AR-low
Fraction of cells in group (%)
20 40 60 80 100
Mean expression in group
0 5 10

Fig. S10 Gene expression profiles for AR-positive and AR-low transformed LumAcinar cells. **A** Dot plot of the 90 most differently expressed genes in LumAcinar cells of the AR-low subgroup, corresponding to the AR-low signature. **B** Dot plot of the Hallmark Androgen Response Signature genes. **C** PHATE plot of the mean expression of the Hallmark Androgen Response gene set. **D** Dot plot of the AR-positive signature from [33]. Both signatures depict a broad negative correlation with the AR-low signature genes in (A).


A
PHATE 2
PHATE 1
Normal
Intermediate
Hyperplasia
ERG-positive
AR-positive
AR-low
ASCL2-positive
ONECUT2-positive


B
Normal
Intermediate
AR-positive
ERG-positive
Hyperplasia
AR-low
ASCL2-positive
POU5F1
SHH
TNF
POU2F3
ASCL2
UCHL1
ENO2
VGF
MSX2
ETV7
Fraction of cells in group (%)
20
40
60
80
Mean expression in group
0.00
0.25
0.50

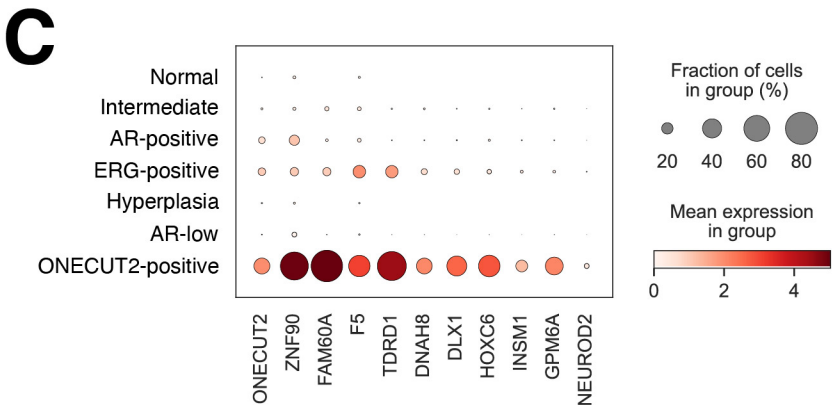
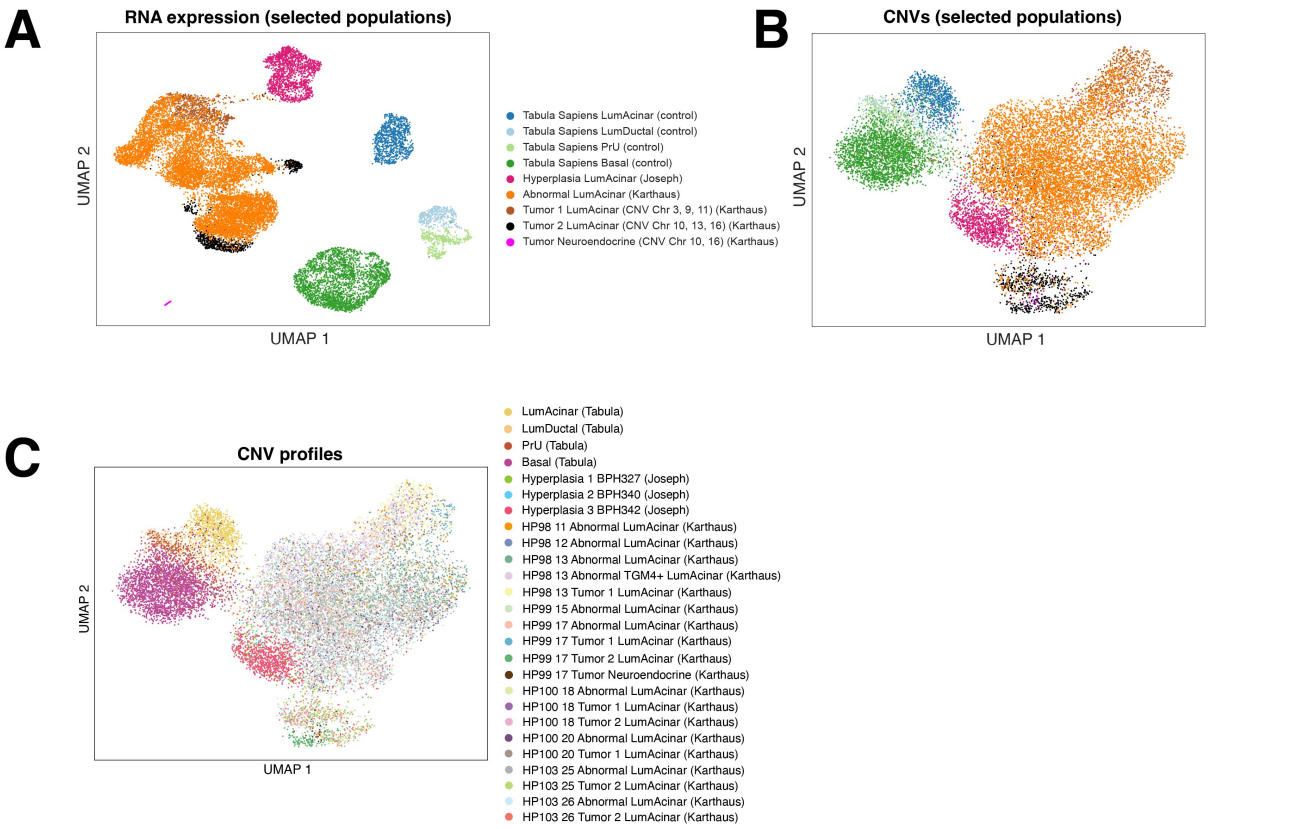
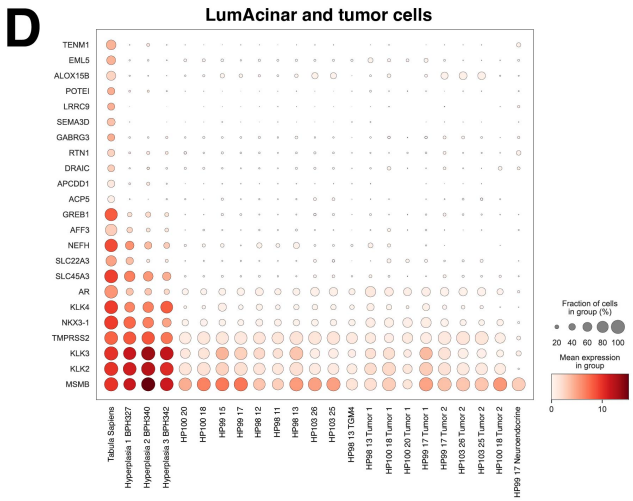
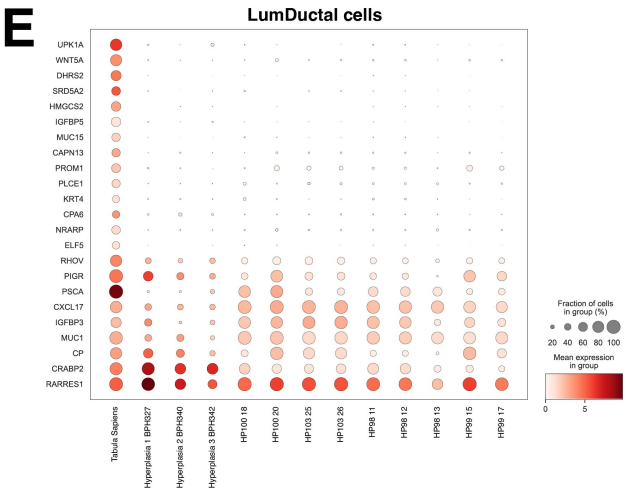

C
Normal
Intermediate
AR-positive
ERG-positive
Hyperplasia
AR-low
ONECUT2-positive
ONECUT2
ZNF90
FAM60A
F5
TDRD1
DNAH8
DLX1
HOXC6
INSM1
GPM6A
NEUROD2
Fraction of cells in group (%)
20
40
60
80
Mean expression in group
0
2
4

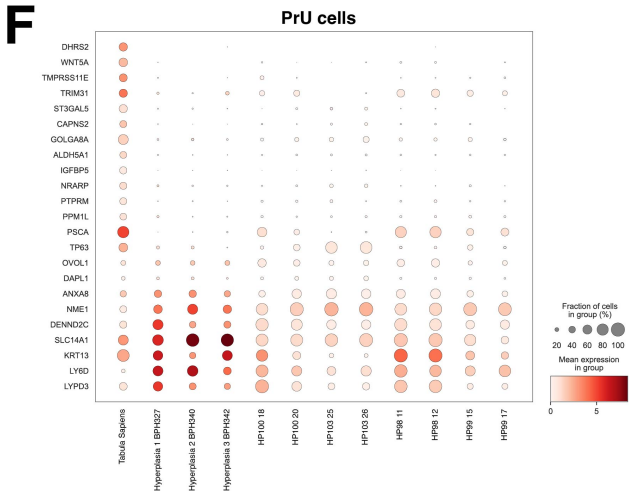
Fig. S11 Markers of neuroendocrine differentiation in tumor and abnormal acinar cells. **A** PHATE plot highlighting the location of two distinct subsets of cells expressing neuroendocrine markers were identified through manual curation based on collective expression of genes associated with neuroendocrine differentiation within the entire group. The ASCL2-positive subset predominantly overlaps with the AR-low population, and the ONECUT2-positive subset aligns with the AR-positive and ERG-positive populations. **B, C** Dot plots depicting selected genes that characterize the ASCL2-positive (B) and ONECUT2-positive (C) subsets.

Figure S12


A
RNA expression (selected populations)
UMAP 2
UMAP 1
Tabula Sapiens LumAcinar (control)
Tabula Sapiens LumDuctal (control)
Tabula Sapiens PrU (control)
Tabula Sapiens Basal (control)
Hyperplasia LumAcinar (Joseph)
Abnormal LumAcinar (Karthaus)
Tumor 1 LumAcinar (CNV Chr 3, 9, 11) (Karthaus)
Tumor 2 LumAcinar (CNV Chr 10, 13, 16) (Karthaus)
Tumor Neuroendocrine (CNV Chr 10, 16) (Karthaus)
B
CNVs (selected populations)
UMAP 2
UMAP 1
C
CNV profiles
UMAP 2
UMAP 1
LumAcinar (Tabula)
LumDuctal (Tabula)
PrU (Tabula)
Basal (Tabula)
Hyperplasia 1 BPH327 (Joseph)
Hyperplasia 2 BPH340 (Joseph)
Hyperplasia 3 BPH342 (Joseph)
HP98 11 Abnormal LumAcinar (Karthaus)
HP98 12 Abnormal LumAcinar (Karthaus)
HP98 13 Abnormal LumAcinar (Karthaus)
HP98 13 Abnormal TGM4+ LumAcinar (Karthaus)
HP98 13 Tumor 1 LumAcinar (Karthaus)
HP99 15 Abnormal LumAcinar (Karthaus)
HP99 17 Abnormal LumAcinar (Karthaus)
HP99 17 Tumor 1 LumAcinar (Karthaus)
HP99 17 Tumor 2 LumAcinar (Karthaus)
HP99 17 Tumor Neuroendocrine (Karthaus)
HP100 18 Abnormal LumAcinar (Karthaus)
HP100 18 Tumor 1 LumAcinar (Karthaus)
HP100 18 Tumor 2 LumAcinar (Karthaus)
HP100 20 Abnormal LumAcinar (Karthaus)
HP100 20 Tumor 1 LumAcinar (Karthaus)
HP103 25 Abnormal LumAcinar (Karthaus)
HP103 25 Tumor 2 LumAcinar (Karthaus)
HP103 26 Abnormal LumAcinar (Karthaus)
HP103 26 Tumor 2 LumAcinar (Karthaus)


D
LumAcinar and tumor cells
TENM1
EML5
ALOX15B
POTEI
LRRC9
SEMA3D
GABRG3
RTN1
DRAIC
APCDD1
ACP5
GREB1
AFF3
NEFH
SLC22A3
SLC45A3
AR
KLK4
NKX3-1
TMPRSS2
KLK3
KLK2
MSMB
Tabula Sapiens
Hyperplasia 1 BPH327
Hyperplasia 2 BPH340
Hyperplasia 3 BPH342
HP100 20
HP100 18
HP99 15
HP99 17
HP98 12
HP98 11
HP98 13
HP103 26
HP103 25
HP98 13 TGM4
HP98 13 Tumor 1
HP100 18 Tumor 1
HP100 20 Tumor 1
HP99 17 Tumor 1
HP99 17 Tumor 2
HP103 26 Tumor 2
HP103 25 Tumor 2
HP100 18 Tumor 2
HP99 17 Neuroendocrine
Fraction of cells in group (%)
20 40 60 80 100
Mean expression in group
0
10


E
LumDuctal cells
UPK1A
WNT5A
DHRS2
SRD5A2
HMGCS2
IGFBP5
MUC15
CAPN13
PROM1
PLCE1
KRT4
CPA6
NRARP
ELF5
RHOV
PIGR
PSCA
CXCL17
IGFBP3
MUC1
CP
CRABP2
RARRES1
Tabula Sapiens
Hyperplasia 1 BPH327
Hyperplasia 2 BPH340
Hyperplasia 3 BPH342
HP100 18
HP100 20
HP103 25
HP103 26
HP98 11
HP98 12
HP98 13
HP99 15
HP99 17
Fraction of cells in group (%)
20 40 60 80 100
Mean expression in group
0
5


F
PrU cells
DHRS2
WNT5A
TMPRSS11E
TRIM31
ST3GAL5
CAPNS2
GOLGA8A
ALDH5A1
IGFBP5
NRARP
PTPRM
PPM1L
PSCA
TP63
OVOL1
DAPL1
ANXA8
NME1
DENND2C
SLC14A1
KRT13
LY6D
LYPD3
Tabula Sapiens
Hyperplasia 1 BPH327
Hyperplasia 2 BPH340
Hyperplasia 3 BPH342
HP100 18
HP100 20
HP103 25
HP103 26
HP98 11
HP98 12
HP99 15
HP99 17
Fraction of cells in group (%)
20 40 60 80 100
Mean expression in group
0
5

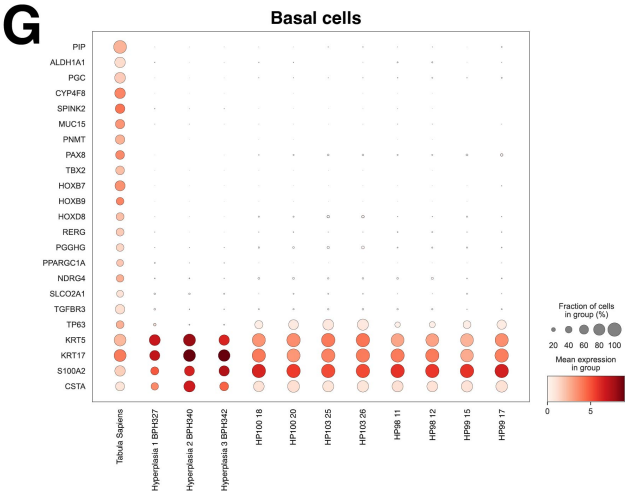

G
Basal cells
PIP
ALDH1A1
PGC
CYP4F8
SPINK2
MUC15
PNMT
PAX8
TBX2
HOXB7
HOXB9
HOXD8
RERG
PGGHG
PPARGC1A
NDRG4
SLCO2A1
TGFBR3
TP63
KRT5
KRT17
S100A2
CSTA
Tabula Sapiens
Hyperplasia 1 BPH327
Hyperplasia 2 BPH340
Hyperplasia 3 BPH342
HP100 18
HP100 20
HP103 25
HP103 26
HP98 11
HP98 12
HP99 15
HP99 17
Fraction of cells in group (%)
20 40 60 80 100
Mean expression in group
0
5

Fig. S12 Meta-analysis of human prostate scRNA-seq data across disease states reveals altered gene expression profiles and a LumAcinar bias of CNVs. **A,B** UMAP plots of RNA expression (A) and CNV profiles (B) of epithelial populations in 2 normal prostates [13], 3 patients with BPH [7], and 4 patients with treatment-naive PCa [6]. To identify tumor populations, inferCNV was used to detect CNVs across the disease samples, normalized against all epithelial populations of the normal prostate from Tabula Sapiens [13], and clusters with CNVs were labeled as definitive tumors. **C** UMAP plot of CNV profiles for each acinar cluster shows that hyperplastic cells have similar CNV profiles to abnormal acinar cells while definitive tumor cells are distinct. **D-G** Dot plots of differentially expressed genes, for the LumAcinar and NE populations (D), LumDuctal (E), PrU (F), and Basal cells (G) across these datasets. For LumAcinar and NE cells in which tumor populations were detected, populations are listed in order of CNV enrichment. Population names match the names used in the source datasets, including spatial regions for each sample from [6]; additional classification was added in (D) to distinguish tumor acinar cells from abnormal acinar cells.

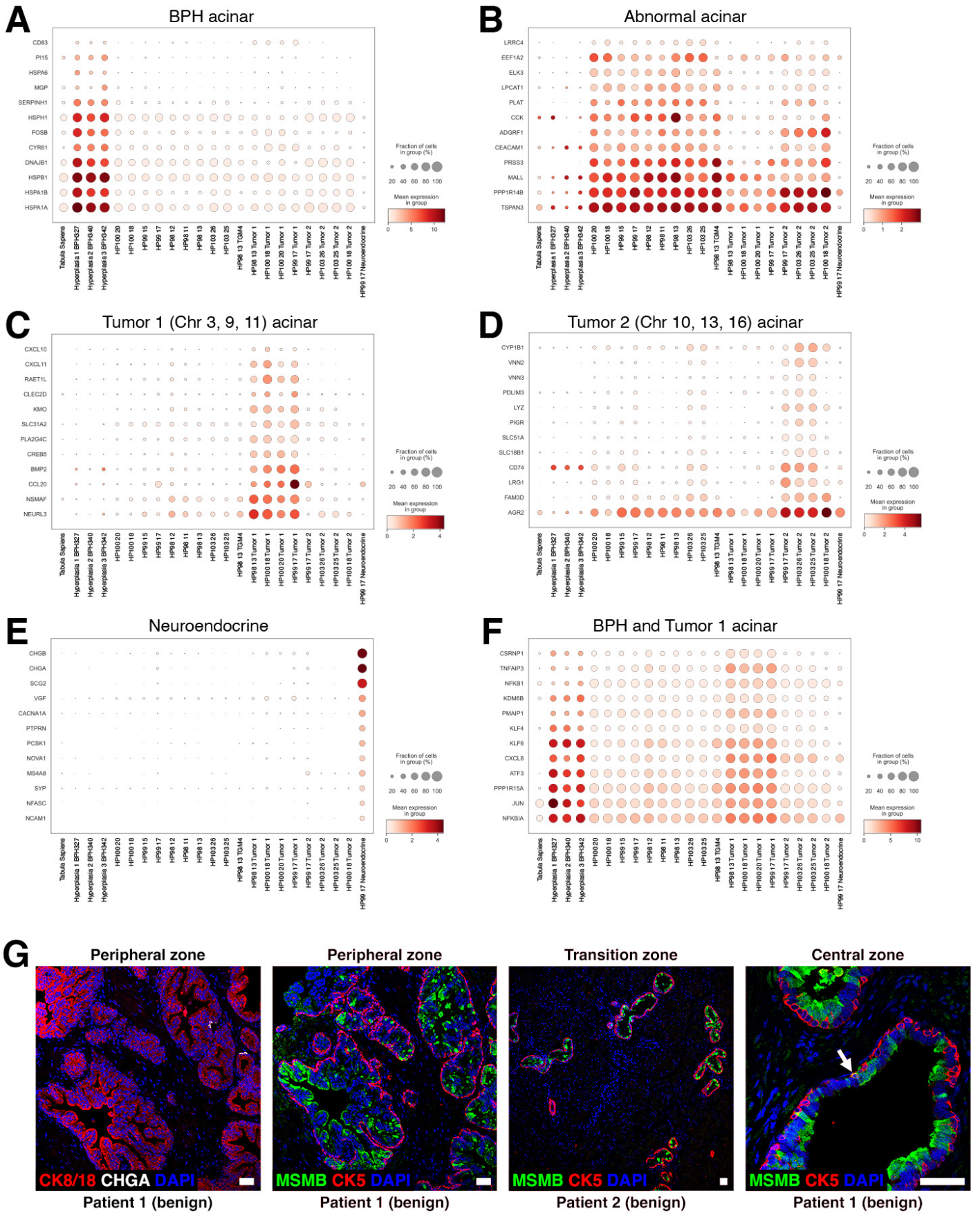

A
BPH acinar
CD83
PI15
HSPA6
MGP
SERPINH1
HSPH1
FOSB
CYR61
DNAJB1
HSPB1
HSPA1B
HSPA1A
B
Abnormal acinar
LRRC4
EEF1A2
ELK3
LPCAT1
PLAT
CCK
ADGRF1
CEACAM1
PRSS3
MALL
PPP1R14B
TSPAN3
C
Tumor 1 (Chr 3, 9, 11) acinar
CXCL10
CXCL11
RAET1L
CLEC2D
KMO
SLC31A2
PLA2G4C
CREB5
BMP2
CCL20
NSMAF
NEURL3
D
Tumor 2 (Chr 10, 13, 16) acinar
CYP1B1
VNN2
VNN3
PDLIM3
LYZ
PIGR
SLC51A
SLC18B1
CD74
LRG1
FAM3D
AGR2
E
Neuroendocrine
CHGB
CHGA
SCG2
VGF
CACNA1A
PTPRN
PCSK1
NOVA1
MS4A8
SYP
NFASC
NCAM1
F
BPH and Tumor 1 acinar
CSRNP1
TNFAIP3
NFKB1
KDM6B
PMAIP1
KLF4
KLF6
CXCL8
ATF3
PPP1R15A
JUN
NFKBIA
Fraction of cells in group (%)
Mean expression in group
Tabula Sapiens
Hyperplasia 1 BPH327
Hyperplasia 2 BPH340
Hyperplasia 3 BPH342
HP100 20
HP100 18
HP99 15
HP99 17
HP98 12
HP98 11
HP98 13
HP103 26
HP103 25
HP98 13 TGM4
HP98 13 Tumor 1
HP100 18 Tumor 1
HP100 20 Tumor 1
HP99 17 Tumor 1
HP99 17 Tumor 2
HP103 26 Tumor 2
HP103 25 Tumor 2
HP100 18 Tumor 2
HP99 17 Neuroendocrine
G
Peripheral zone
CK8/18 CHGA DAPI
Patient 1 (benign)
Peripheral zone
MSMB CK5 DAPI
Patient 1 (benign)
Transition zone
MSMB CK5 DAPI
Patient 2 (benign)
Central zone
MSMB CK5 DAPI
Patient 1 (benign)

Fig. S13 Marker expression for the LumAcinar and NE populations across disease states. **A-F** Dot plots of specific markers for BPH acinar cells (A), abnormal acinar cells (B), Tumor CNV group 1 acinar cells (C), Tumor CNV group 2 acinar cells (D), Neuroendocrine tumor cells (E), and both BPH and Tumor 1 acinar cells (F). Note that BPH acinar cells uniquely express genes for many heat shock proteins, and that abnormal acinar cells generally share several changes in gene expression with definitive tumor acinar cells, even when these cells were isolated from different regions of the prostate. For example, samples 11, 12, and 13 of HP98 have abnormal acinar cells though definitive tumor cells are only found in sample 13, and samples 15 and 17 of patient HP99 have abnormal acinar cells though definitive tumor cells are found in only sample 17. **G** Immunofluorescence analysis of cell type and LumAcinar markers indicates that transcriptomic changes in acinar cells can be broadly observed, as shown by maintenance of luminal CK8/18 expression but loss of acinar-specific MSMB expression. *(Left)* MSMB is expressed in CK8/18-positive LumAcinar cells in the peripheral zone, and is absent where Basal cells are still intact, independent of proximity to NE cells. *(Middle right)* Reduced MSMB expression in LumAcinar cells can also be observed in the transition zone, which is less frequently a site of prostate cancer. *(Right)* Reduced MSMB expression can also be observed in the central zone, which has the lowest rate of prostate cancer. A rare example of MSMB expression in a CK5+MSMB+ intermediate-like cell is highlighted (white arrow). Notably, basal cells are intact in all samples, indicating that these changes in LumAcinar gene expression are detected in benign prostate tissue regions.

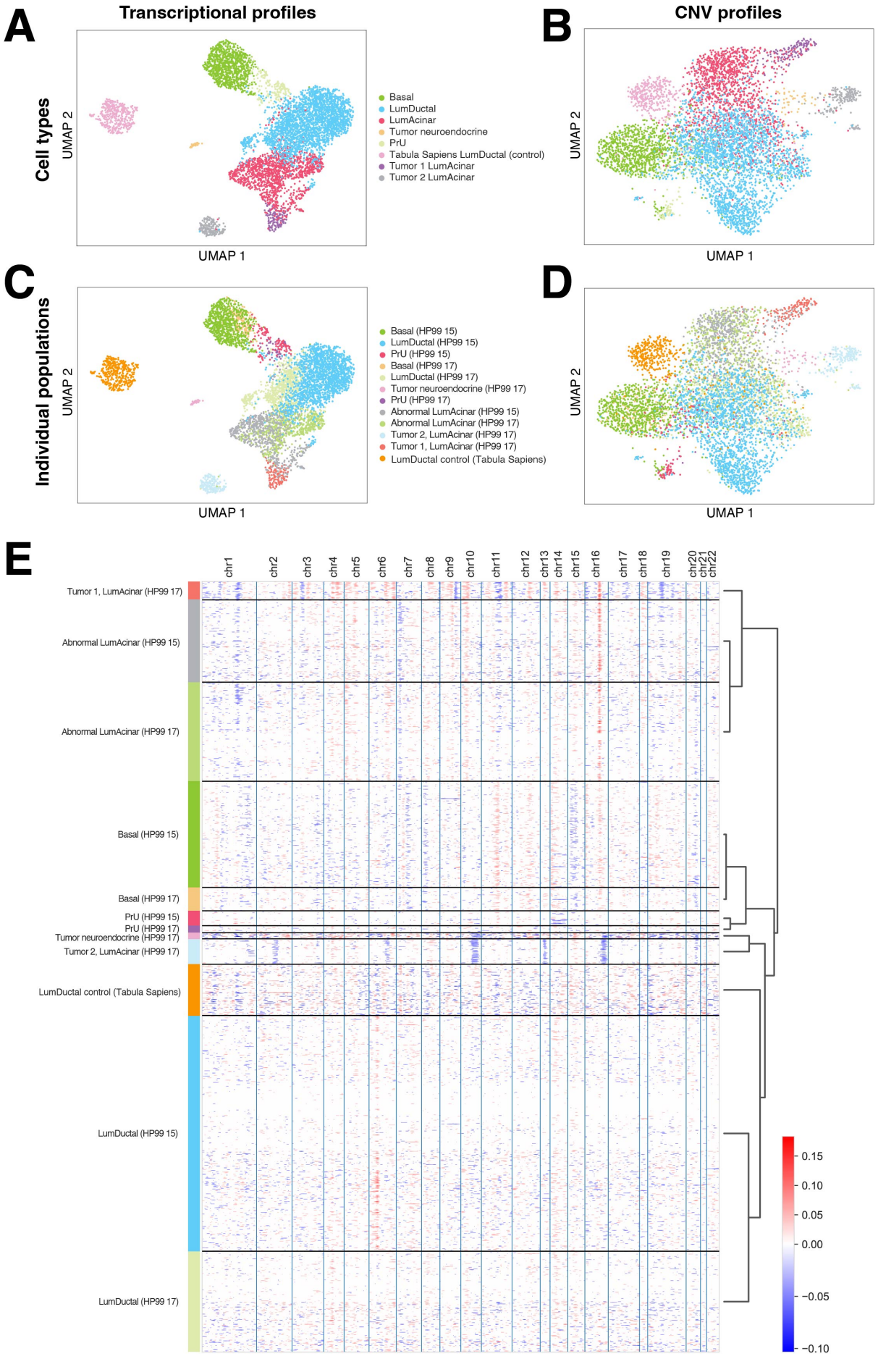

A
Transcriptional profiles
Cell types
UMAP 2
UMAP 1
Basal
LumDuctal
LumAcinar
Tumor neuroendocrine
PrU
Tabula Sapiens LumDuctal (control)
Tumor 1 LumAcinar
Tumor 2 LumAcinar
B
CNV profiles
UMAP 2
UMAP 1
C
Individual populations
UMAP 2
UMAP 1
Basal (HP99 15)
LumDuctal (HP99 15)
PrU (HP99 15)
Basal (HP99 17)
LumDuctal (HP99 17)
Tumor neuroendocrine (HP99 17)
PrU (HP99 17)
Abnormal LumAcinar (HP99 15)
Abnormal LumAcinar (HP99 17)
Tumor 2, LumAcinar (HP99 17)
Tumor 1, LumAcinar (HP99 17)
LumDuctal control (Tabula Sapiens)
D
UMAP 2
UMAP 1
E
chr1
chr2
chr3
chr4
chr5
chr6
chr7
chr8
chr9
chr10
chr11
chr12
chr13
chr14
chr15
chr16
chr17
chr18
chr19
chr20
chr21
chr22
Tumor 1, LumAcinar (HP99 17)
Abnormal LumAcinar (HP99 15)
Abnormal LumAcinar (HP99 17)
Basal (HP99 15)
Basal (HP99 17)
PrU (HP99 15)
PrU (HP99 17)
Tumor neuroendocrine (HP99 17)
Tumor 2, LumAcinar (HP99 17)
LumDuctal control (Tabula Sapiens)
LumDuctal (HP99 15)
LumDuctal (HP99 17)
0.15
0.10
0.05
0.00
−0.05
−0.10

Fig. S14 InferCNV methodology to distinguish tumor and abnormal cell populations. **A-D** UMAP plots of gene expression (A,C) and CNVs (B,D) in prostate cell populations from patient HP99 [6]. (A,B) show general cell populations and the Tabula Sapiens Luminal Ductal cells (used for CNV normalization), and (C, D) show specific populations detected by the RMT pipeline. **E** Heatmap of detected CNVs across sections 15 and 17 taken from HP99, as well as the control LumDuctal cells [13] used for normalization. A population was determined to be a definitive tumor population if over 25% of the cells within the population were enriched for the same CNV over a control. HP99 has a Tumor 1 group population (CNVs on chromosomes 3, 9 & 11) and additional patient-specific CNVs, a Tumor 2 group population (CNVs on chromosomes 10, 13 & 16) and additional CNVs, and a NE Tumor population (CNVs on chromosomes 10 & 16). This *de novo* NE tumor population shares several CNVs with an acinar Tumor 2 population from the same region. Definitive tumor populations for HP99 were all detected in section 17, though transcriptomic changes could be detected throughout populations in both sections 15 and 17, which have unique cell type compositions.

Figure S15

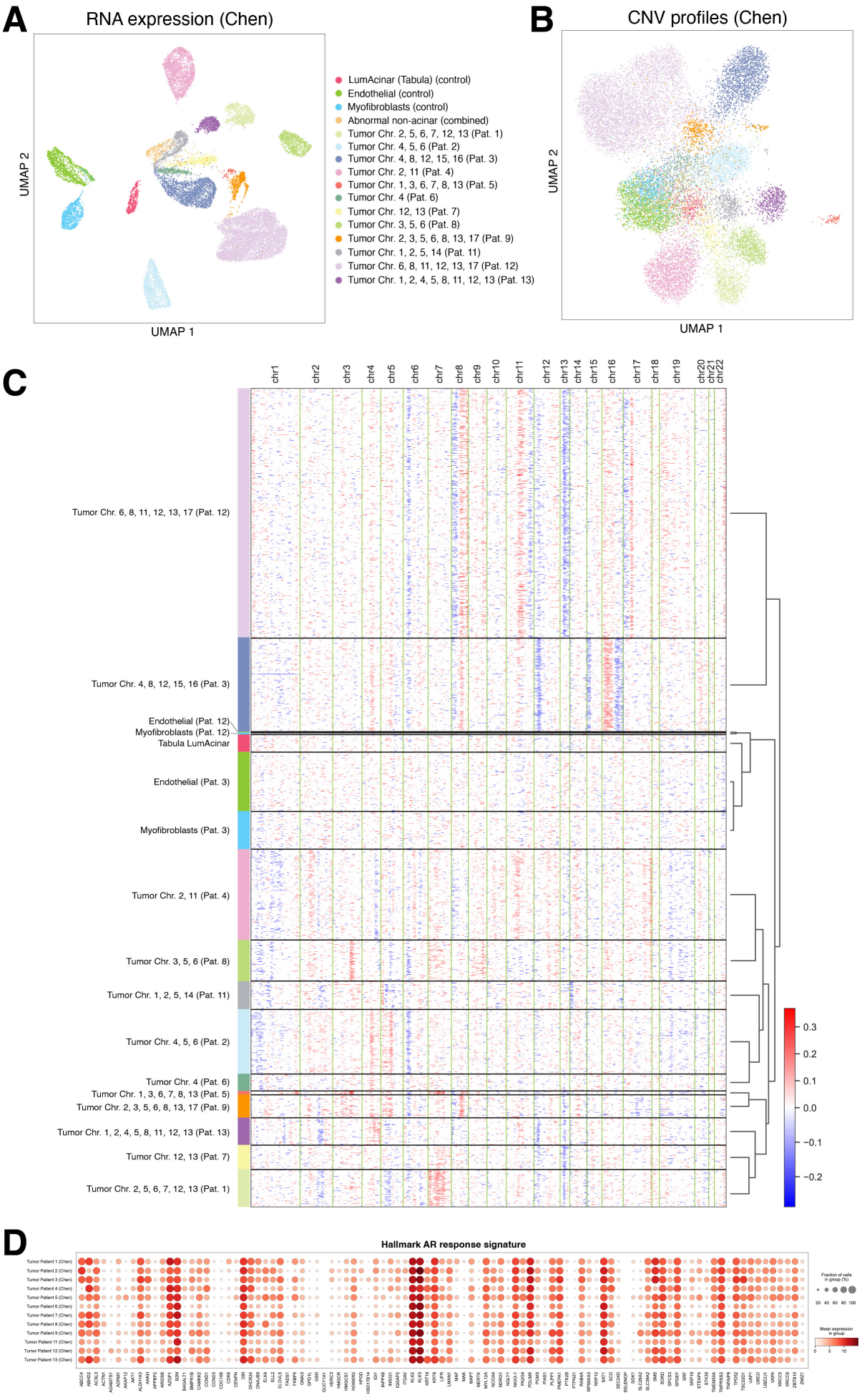
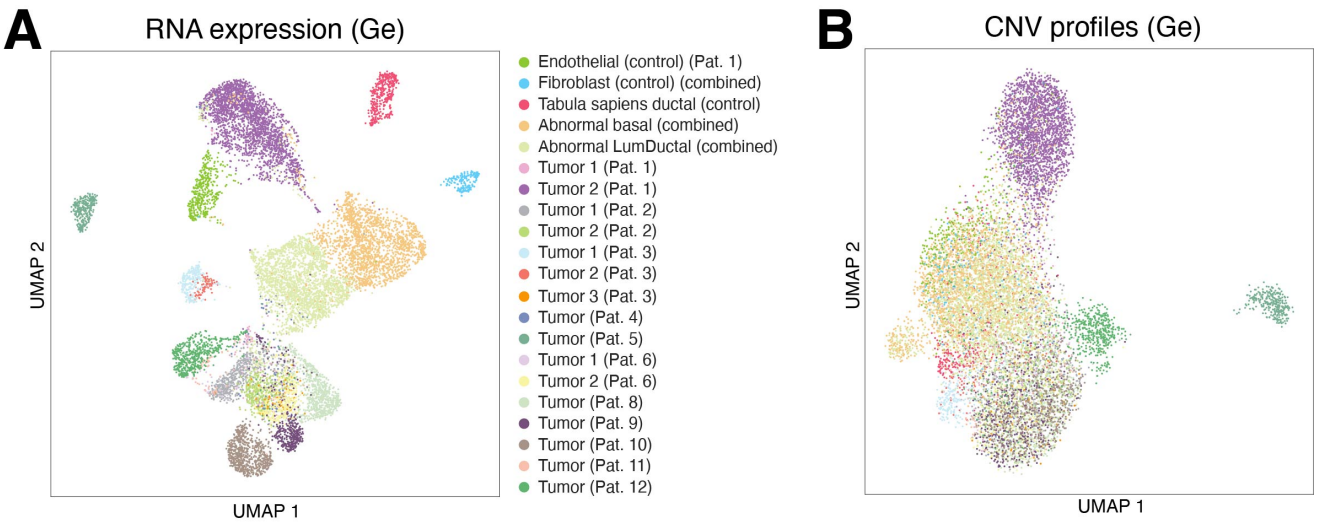

A
RNA expression (Chen)
UMAP 2
UMAP 1
LumAcinar (Tabula) (control)
Endothelial (control)
Myofibroblasts (control)
Abnormal non-acinar (combined)
Tumor Chr. 2, 5, 6, 7, 12, 13 (Pat. 1)
Tumor Chr. 4, 5, 6 (Pat. 2)
Tumor Chr. 4, 8, 12, 15, 16 (Pat. 3)
Tumor Chr. 2, 11 (Pat. 4)
Tumor Chr. 1, 3, 6, 7, 8, 13 (Pat. 5)
Tumor Chr. 4 (Pat. 6)
Tumor Chr. 12, 13 (Pat. 7)
Tumor Chr. 3, 5, 6 (Pat. 8)
Tumor Chr. 2, 3, 5, 6, 8, 13, 17 (Pat. 9)
Tumor Chr. 1, 2, 5, 14 (Pat. 11)
Tumor Chr. 6, 8, 11, 12, 13, 17 (Pat. 12)
Tumor Chr. 1, 2, 4, 5, 8, 11, 12, 13 (Pat. 13)
B
CNV profiles (Chen)
UMAP 2
UMAP 1
C
chr1
chr2
chr3
chr4
chr5
chr6
chr7
chr8
chr9
chr10
chr11
chr12
chr13
chr14
chr15
chr16
chr17
chr18
chr19
chr20
chr21
chr22
Tumor Chr. 6, 8, 11, 12, 13, 17 (Pat. 12)
Tumor Chr. 4, 8, 12, 15, 16 (Pat. 3)
Endothelial (Pat. 12)
Myofibroblasts (Pat. 12)
Tabula LumAcinar
Endothelial (Pat. 3)
Myofibroblasts (Pat. 3)
Tumor Chr. 2, 11 (Pat. 4)
Tumor Chr. 3, 5, 6 (Pat. 8)
Tumor Chr. 1, 2, 5, 14 (Pat. 11)
Tumor Chr. 4, 5, 6 (Pat. 2)
Tumor Chr. 4 (Pat. 6)
Tumor Chr. 1, 3, 6, 7, 8, 13 (Pat. 5)
Tumor Chr. 2, 3, 5, 6, 8, 13, 17 (Pat. 9)
Tumor Chr. 1, 2, 4, 5, 8, 11, 12, 13 (Pat. 13)
Tumor Chr. 12, 13 (Pat. 7)
Tumor Chr. 2, 5, 6, 7, 12, 13 (Pat. 1)
0.3
0.2
0.1
0.0
−0.1
−0.2
D
Hallmark AR response signature
Tumor Patient 1 (Chen)
Tumor Patient 2 (Chen)
Tumor Patient 3 (Chen)
Tumor Patient 4 (Chen)
Tumor Patient 5 (Chen)
Tumor Patient 6 (Chen)
Tumor Patient 7 (Chen)
Tumor Patient 8 (Chen)
Tumor Patient 9 (Chen)
Tumor Patient 11 (Chen)
Tumor Patient 12 (Chen)
Tumor Patient 13 (Chen)
Fraction of cells in group (%)
20 40 60 80 100
Mean expression in group
0 5 10

Fig. S15 InferCNV analysis of prostate tumors from the Chen cohort. **A,B** UMAP plots of gene expression (A) and CNVs (B) for the cell populations from [17] and a normalization control from [13]. **C** Heatmap of CNV enrichment across pooled patient samples and control populations. Data were pooled and CNVs were normalized as a group to both internal mesenchymal populations and external normal LumAcinar cells from Tabula Sapiens. All definitive tumor populations are LumAcinar, and this cohort has completely variable CNV patterns between patients. **D** Dot plot of expression of Hallmark AR response genes in individual samples from the Chen cohort. These authors stated that one of the patients in their cohort underwent androgen-deprivation therapy, but it is unclear from their annotation whether this patient sample was sequenced or included in their analysis. However, since each of these samples displays robust expression of AR response genes, we conclude that all of the samples that we analyzed were from hormonally-intact patients.

Figure S16


A
RNA expression (Ge)
UMAP 2
UMAP 1
Endothelial (control) (Pat. 1)
Fibroblast (control) (combined)
Tabula sapiens ductal (control)
Abnormal basal (combined)
Abnormal LumDuctal (combined)
Tumor 1 (Pat. 1)
Tumor 2 (Pat. 1)
Tumor 1 (Pat. 2)
Tumor 2 (Pat. 2)
Tumor 1 (Pat. 3)
Tumor 2 (Pat. 3)
Tumor 3 (Pat. 3)
Tumor (Pat. 4)
Tumor (Pat. 5)
Tumor 1 (Pat. 6)
Tumor 2 (Pat. 6)
Tumor (Pat. 8)
Tumor (Pat. 9)
Tumor (Pat. 10)
Tumor (Pat. 11)
Tumor (Pat. 12)
B
CNV profiles (Ge)
UMAP 2
UMAP 1

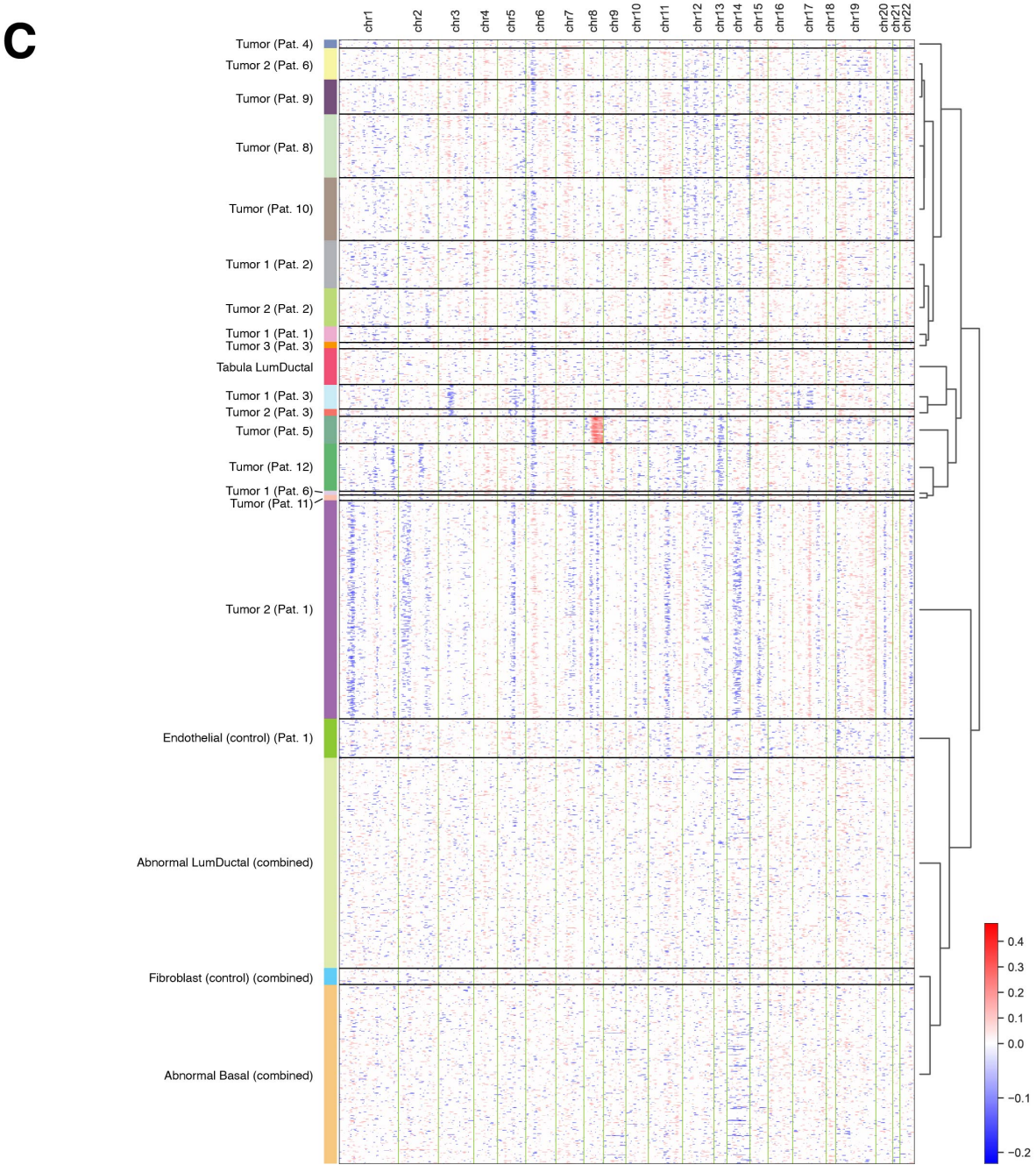
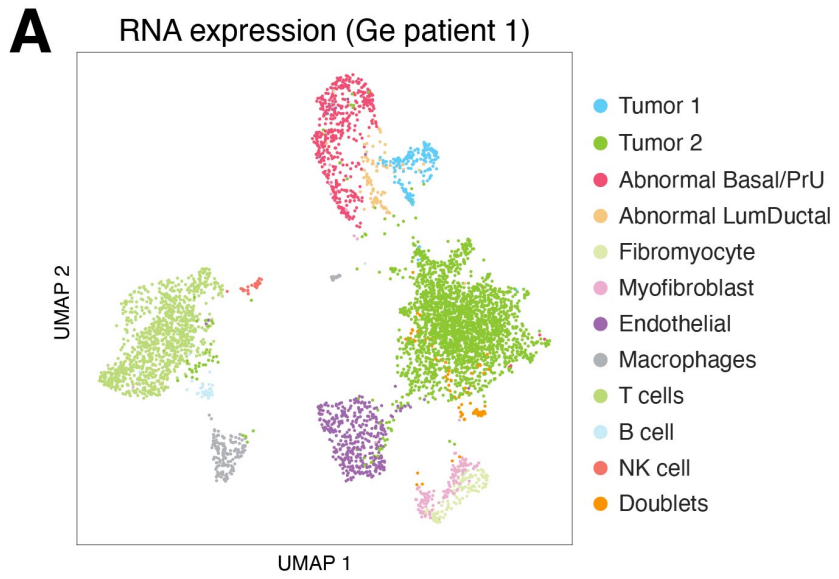

C
chr1
chr2
chr3
chr4
chr5
chr6
chr7
chr8
chr9
chr10
chr11
chr12
chr13
chr14
chr15
chr16
chr17
chr18
chr19
chr20
chr21
chr22
Tumor (Pat. 4)
Tumor 2 (Pat. 6)
Tumor (Pat. 9)
Tumor (Pat. 8)
Tumor (Pat. 10)
Tumor 1 (Pat. 2)
Tumor 2 (Pat. 2)
Tumor 1 (Pat. 1)
Tumor 3 (Pat. 3)
Tabula LumDuctal
Tumor 1 (Pat. 3)
Tumor 2 (Pat. 3)
Tumor (Pat. 5)
Tumor (Pat. 12)
Tumor 1 (Pat. 6)
Tumor (Pat. 11)
Tumor 2 (Pat. 1)
Endothelial (control) (Pat. 1)
Abnormal LumDuctal (combined)
Fibroblast (control) (combined)
Abnormal Basal (combined)
0.4
0.3
0.2
0.1
0.0
−0.1
−0.2

Fig. S16 InferCNV analysis of prostate tumors from the Ge cohort. **A,B** UMAP plots of gene expression (A) and CNVs (B) for the cell populations from 12 PCa patients [16] and a normalization control from [13]. **C** Heatmap of CNV enrichment across the patient samples. Data were pooled and CNVs were normalized as a group to multiple internal mesenchymal and epithelial populations as well as external normal LumAcinar cells from Tabula Sapiens. This cohort has CNV patterns that are observed in several patients; further, many patients had multiple tumor populations with distinct sets of CNVs, which we have labeled numerically. All definitive tumor populations are LumAcinar.

Figure S17


A
RNA expression (Ge patient 1)
UMAP 2
UMAP 1
Tumor 1
Tumor 2
Abnormal Basal/PrU
Abnormal LumDuctal
Fibromyocyte
Myofibroblast
Endothelial
Macrophages
T cells
B cell
NK cell
Doublets

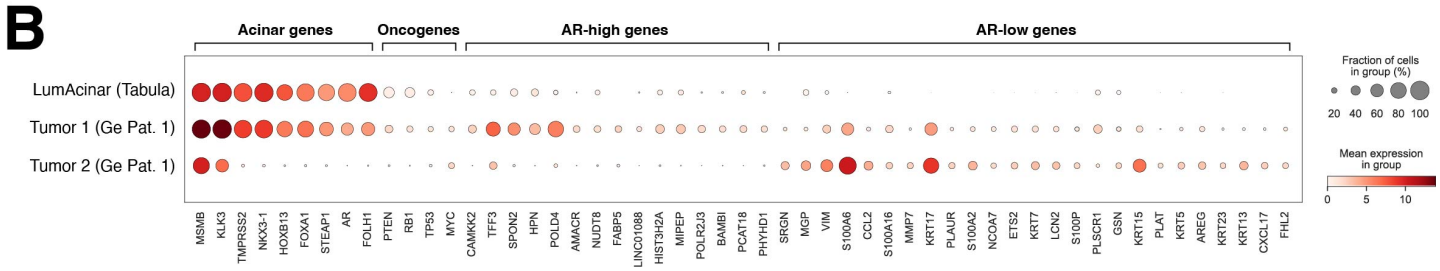

B
Acinar genes
Oncogenes
AR-high genes
AR-low genes
LumAcinar (Tabula)
Tumor 1 (Ge Pat. 1)
Tumor 2 (Ge Pat. 1)
MSMB KLK3 TMPRSS2 NKX3-1 HOXB13 FOXA1 STEAP1 AR FOLH1
PTEN RB1 TP53 MYC
CAMKK2 TFF3 SPON2 HPN POLD4 AMACR NUDT8 FABP5 LINC01088 HIST3H2A MIPEP POLR2J3 BAMBI PCAT18 PHYHD1
SRGN MGP VIM S100A6 CCL2 S100A16 MMP7 KRT17 PLAUR S100A2 NCOA7 ETS2 KRT7 LCN2 S100P PLSCR1 GSN KRT15 PLAT KRT5 AREG KRT23 KRT13 CXCL17 FHL2
Fraction of cells in group (%)
20 40 60 80 100
Mean expression in group
0 5 10

Fig. S17 Full InferCNV analysis of the patient 1 tumor from the Ge cohort. **A** UMAP plots of gene expression for cell populations from patient 1 [16], showing two distinct tumor populations and multiple abnormal epithelial cell types. **B** Dot plot of the normal LumAcinar cells (Tabula), and Patient 1 LumAcinar Tumor 1 and Tumor 2 cells, showing that Patient 1 had both AR-positive and AR-low tumor cells concurrently.