ERG-driven prostate cancer emerges from basal-luminal hybrid cells

Weiran Feng[1,10], Erik Ladewig[2,10], Nazifa Salsabeel[1,11], Huiyong Zhao[3,11], Young Sun Lee[1], Anuradha Gopalan[4], Hanzhi Luo[5], Wenfei Kang[6], Ning Fan[6], Eric Rosiek[6], Ignas Masilionis[2], Ronan Chaligné[2], Elisa de Stanchina[3], Yu Chen[1], Brett S. Carver[1,7,8], Christina S Leslie[2], Charles L. Sawyers[1,9]*

[1]Human Oncology and Pathogenesis Program, Memorial Sloan Kettering Cancer Center; New York, NY 10065, USA.

[2]Computational and Systems Biology Program, Memorial Sloan Kettering Cancer Center; New York, NY 10065, USA.

[3]Antitumor Assessment Core Facility, Memorial Sloan Kettering Cancer Center; New York, NY 10065, USA.

[4]Department of Pathology, Memorial Sloan Kettering Cancer Center; New York, NY 10065, USA.

[5]Molecular Pharmacology Program, Memorial Sloan Kettering Cancer Center; New York, NY 10065, USA.

[6]Molecular Cytology Core Facility, Memorial Sloan Kettering Cancer Center; New York, NY 10065, USA.

[7]Department of Surgery, Memorial Sloan Kettering Cancer Center; New York, NY 10065, USA.

[8]Division of Urology, Memorial Sloan Kettering Cancer Center; New York, NY 10065, USA.

[9]Howard Hughes Medical Institute, Memorial Sloan Kettering Cancer Center; New York, NY 10065, USA.

[10]These authors contributed equally to this work.

[11]These authors contributed equally to this work.

*Corresponding author. Email: sawyersc@mskcc.org

Abstract:

To gain insight into how ERG translocations cause prostate cancer, we performed single cell transcriptional profiling of an autochthonous mouse model at an early stage of disease initiation. Despite broad expression of ERG in prostate epithelial cells, we observed enhanced proliferation primarily in an intermediate subpopulation with a basal identity based on transcriptomics but with luminal morphology and cytokeratin expression. Through a series of lineage tracing and primary prostate epithelial transplantation experiments, we find that tumor initiating activity resides in a subpopulation of basal cells that also express luminal genes such as *Tmprss2* and *Nkx3.1*, but not in the larger population of classical luminal cells. Upon ERG activation, these basal-luminal hybrid cells expand into a highly proliferative population of intermediate basal-luminal identity then transition to fully differentiated luminal cells that reflect the lineage of ERG-positive cancers. These findings help resolve a preexisting debate about the cell of origin of ERG-driven prostate cancer and narrow the focus for future mechanistic studies to a basal-luminal subpopulation of tumor initiating cells.

Main Text:

Translocations of the ETS family transcription factor ERG are found in nearly 150,000 newly diagnosed prostate cancers (PCas) each year in the United States (*1, 2*), underscoring the importance of defining the mechanism by which ERG promotes prostate epithelial transformation. A consequence of nearly all ERG translocations is aberrant expression of ERG in prostate luminal cells under transcriptional control of TMPRSS2. Multiple lines of evidence implicate ERG translocations as an initiating event in prostate cancer. ERG expression is seen in prostatic intraepithelial neoplasia (PIN) and proliferative inflammatory atrophy; expression in cancers is typically uniform; and, in the context of PI3K pathway activation, ERG is sufficient to induce cancers in mice (*3-10*).

The mechanism by which ERG initiates prostate cancer remains unclear despite extensive research (*5, 11-18*). When expressed in primary prostate epithelial cells (in mice or organoids), ERG binds multiple ETS sites throughout the genome but only promotes downstream transcriptional changes in a setting of PI3K activation (*5*), which occurs through PTEN loss or PIK3CA family gene mutations in a large fraction of patients with ERG translocations (*6, 7, 19*). Despite these clues, the downstream transcriptional changes responsible for oncogenicity are largely unknown. Bulk RNA-seq analysis of prostate tumors from an autochthonous ERG-driven mouse model and of ERG+ organoids has shown that ERG enhances androgen receptor transcriptional output and activates a pro-luminal epithelial differentiation program, with consequent loss of basal epithelial cells (*5, 20*). These findings raise an intriguing paradox: how does an initiating oncogene that drives terminal epithelial differentiation also promote oncogenesis?

ERG stimulates proliferation of epithelial cells with an intermediate lineage state

To gain insight into this question, we re-examined transcriptomic changes in the *Rosa26-ERG$^{LSL/LSL}$; Pten$^{flox/flox}$;Pb-Cre4* (*EPC*) mouse model, now using single cell RNA sequencing (scRNA-seq) and at an early timepoint to capture changes associated with cancer initiation. At age 3 months EPC mice have small foci of invasive cancer which subsequently progress to highly penetrant glandular invasion within 6 months (**fig. S1A**). Mice with *Pten* deletion only (*Pten$^{flox/flox}$;Pb-Cre4*, hereafter called *PC*) develop intraductal hyperplasia or PIN early and develop invasive cancer only at advanced age (>9 months) (*5, 21*).

Uniform manifold approximation and projection (UMAP) of all cells from the prostates of *EPC*, *PC* and *WT* mice revealed genotype-specific differences in both epithelial and non-epithelial populations (**Fig. 1A-C, fig. S1B-F).** Focusing initially on the epithelial cells (Epcam+), we noted that the large clusters of L1 luminal cells in *WT* mice (distally located secretory cells that account for >90 percent of luminal cells) (*22-24*) are absent in *EPC* and *PC* mice. In contrast, L2 cells (also known as LumP), which express stem-like markers (Sca1, Trop2) and have increased regenerative potential (*22-24*), are more abundant in *EPC* and *PC* mice (**Fig. 1A-C, fig. S1E-G**) (*25*). More striking, however, is a cluster of highly proliferative cells in *EPC* and *PC* mice that are Ki67+, Top2a^{high}, Pcna^{high} with increased S and G2/M cycle scores, an increased gene signature of stemness and a basal (bas) identity based on expression signature assignments (named Tumor-Bas) (**Fig. 1B-E, fig. S1H**). This basal cluster (Tumor-Bas 10, and the adjacent cluster 1) is distinct from the cluster of normal basal cells in *WT* mice and, in *EPC* mice these cells express ERG (**Fig. 1E**). Cells in Tumor-Bas clusters also express canonical luminal lineage genes (*Krt8/18/19*, *Cd24a*, *Foxa1*) as well as basal genes (*Krt5/14/15/17*, *Trp63*) (**Fig. 1F**), and therefore have an intermediate (IM) identity despite the transcriptomic assignment of a basal identity.

To rule out an artifactual explanation for co-expression of basal and luminal genes (e.g., doublets in single cell analysis), we confirmed the presence of this highly proliferative IM population using flow cytometry. As expected, *WT* mice had two major epithelial populations, single-positive luminal cells (cytokeratin (K)8+/K18+, ~76%, hereafter called SP^{lum}) and single-positive basal cells (K5+/p63+, ~15%, hereafter called SP^{bas}), and a third minor population (~1%) that was double-positive (DP) (**Fig. 1G**). Consistent with the single cell transcriptomic analysis, the SP^{bas} and DP populations from *PC* and *EPC* mice were expanded, with reduced p63 relative to *WT*. EPC mice showed an even greater expansion of DP cells (~27%) compared to PC mice (~8%) (**Fig. 1G-H, fig. S2A-B**). Roughly 40% of the DP cells in *EPC* mice express high levels of Ki67 (**Fig. I, fig. S2C**), confirming the high proliferation rate seen by scRNA-seq. Because the contour plots show that the SP^{bas} and DP quadrants represent a single population, we refer to these as IM cells (**Fig. 1G**). Of note, ERG expression results in a greater luminal shift in this IM population relative to that in *PC* mice, as seen by flow cytometry and by using a probability-based transcriptomic analysis that assigns quantitative basal or luminal lineage scores to each cell (**fig. S3**, see methods). IM cells have a luminal morphology and express luminal markers such as K8, K18 and Ar but also express basal genes such as K5 (**Fig. 1J-M**). In

addition to K5+ IM cells, a large fraction of ERG+ invasive cells with luminal morphology are K5-negative (**Fig. 1J-K**), consistent with the luminal histology of ERG+ human tumors.

Intermediate cells give rise to luminal cells

Having documented the presence of basal-luminal IM cells by scRNA-seq, flow cytometry and histology, we next investigated the lineage relationships between IM cells and the K5-negative luminal cells through EdU (5-ethynyl-2'-deoxyuridine) labeling. As expected, ERG+ IM cells were preferentially labeled over luminal cells after a brief EdU pulse (2 hours). During the 1 week chase, the label shifted toward K8+, K5-negative, ERG+ luminal cells, indicating that ERG+ IM cells give rise to ERG+ luminal cells (**Fig. 2A, fig. S4**). To provide direct visual evidence for this transition, we identified ~300 EdU-labeled cell doublets (dividing cells), of which ~26% showed asymmetric divisions, always with one DP daughter cell paired with a SP^{lum} or SP^{bas} daughter cell. ~80% of these asymmetric divisions consisted of DP-SP^{lum} pairs, providing further evidence that DP cells give rise to SP^{lum} cells (**Fig. 2B-C**).

ERG accelerates luminal output from basal cells

To gain greater clarity into these lineage transitions, we moved from *EPC* mice to a mouse organoid system that allows direct tracing of lineage fate. After establishing primary prostate organoids from $ERG^{LSL/LSL}$; $Pten^{flox/flox}$ mice, we infected the organoids with Cre virus to delete *Pten* and activate *ERG* expression (**fig. S5A-B**), then repeated the EdU pulse chase experiment but now *in vitro*. After the 2 hour pulse, the EdU label was primarily in basal cells (>80%) but shifted to DP and SP^{lum} cells during the chase (35% and 27% respectively). This shift was ERG-dependent (**Fig. 2D-F, fig. S5C**). To address whether ERG induction is sufficient to drive a luminal fate transition in SP^{bas} cells, we generated a dual reporter system in *EP* organoids by knocking fluorescent marker genes into the endogenous *Krt5* and *Krt8* loci (**fig. S6**), thereby providing a platform to isolate live, purified basal or luminal cells for subsequent lineage commitment studies. After confirming the expected localization of the respective K5-targeted (basal) and K8-targeted (luminal) fluorescent signals (**Fig. 2G**), we tracked the lineage fate of sorted basal and luminal cells in organoid culture. ERG-positive basal cells showed a ~5-fold increase in luminal fate transition compared to ERG-negative cells, confirming that ERG

accelerates luminal differentiation from basal cells. However, ERG is not sufficient to sustain luminal fate in sorted luminal cells (**Fig. 2H**).

To summarize, the most highly proliferative epithelial cells during the early stages of tumorigenesis in EPC mice are an ERG+ population of cells with mixed basal-luminal (IM) identity. We also show that ERG accelerates the generation of SP^{lum} luminal cells from SP^{bas} basal cells, a normally restricted process under adult homeostasis but activated in development and injury (*26-30*). This raises the possibility that ERG+ prostate cancers originate in basal cells despite presenting with luminal histology, as has been suggested by previous transplantation studies of human prostate epithelial cells (*31-34*). However, luminal cells can also give rise to adenocarcinomas in transplantation assays and in autochthonous models (*34-40*). Our results documenting an ERG-driven basal-to-luminal transition and expansion of an IM population in ERG+ tumors prompted us to gain further clarity on this issue. To this end, we undertook three complementary experimental strategies in which we selectively activated ERG in either basal or luminal cells then scored tumorigenesis.

Basal cells are preferred cell of origin in transplantation assays

We previously reported a method to interrogate prostate cancer cell of origin by harvesting primary mouse prostate tissue, sorting into luminal and basal subpopulations, genetically manipulating each subpopulation ex vivo, then immediately (same day) reintroducing these cells orthotopically (OT) to score subsequent tumor formation (*40*). Using this approach, luminal cells demonstrated greater tumorigenic potential than basal cells after combined deletion of *Pten* and *Trp53*, confirming results seen previously using luminal or basal-specific Cre drivers (*35*). To address the cell of origin using ERG, we isolated luminal and basal cells from *EP* mice, activated the *EP* genotype by infection with Ad-Cre virus and performed OT implantations (**Fig. 3A**). Post-sort analysis confirmed >85% purity of the sorted populations preimplantation and >90% Cre recombination efficiency (**fig. S7A**). After 5 months, ERG+ cells were detected in both basal and luminal-derived orthografts (**Fig. 3B**) but only the basal-derived orthografts developed invasive adenocarcinomas. Specifically, 6 of 6 mice with basal-derived orthografts developed grossly visible tumors (**fig. S7B-C**) that contained the highly proliferative ERG+ IM cells and ERG+ SP^{lum} cells (**Fig. 3C**). Histologically, the basal-derived orthografts displayed luminal morphology, with evidence of nuclear and nucleolar enlargement, and expressed K8 and K18 (as

well as K5 in IM cells) with reduced p63. In contrast, luminal-derived orthografts were smaller, less penetrant (4 of 6) and displayed benign histology (**Fig. 3D; fig. S7B-D**).

ERG-driven cancers emerge from K5 positive cells

Having implicated basal cells as cell of origin in the OT assay, we turned to conventional lineage tracing approaches to validate our findings in the autochthonous model using tamoxifen-inducible *K5-CreER*^{T2} (basal) and *K8-CreER*^{T2} (luminal) mouse strains. Both are well established tools previously used by others to address lineage questions in breast and prostate tissue (*41-44*). After confirming the lineage fidelity of these strains in prostate by crossing to a YFP reporter strain (**fig. S8**). we generated the relevant *EP* compound mice (*EP*; *K5-CreER*^{T2} and *EP*; *K8-CreER*^{T2}), administered tamoxifen to cohorts of each strain at age 7-9 weeks, then followed mice for disease onset. Surprisingly, all *EP*; *K5-CreER*^{T2} mice had to be sacrificed within 2-3 weeks of Cre induction due to a fulminant skin disease with extensive desquamation. We observed an essentially identical phenotype in a comparable experiment using an independent basal-specific Cre driver strain (*K14-CreER*^{T2}) (**fig. S9**). Because this acute skin phenotype has not been previously reported in *Pten*$^{flox/flox}$ mice, we presume it is a consequence aberrant expression of ERG in K5/14-expressing cells in skin (*45*). Unfortunately, the early death of *EP*; *K5-CreER*^{T2} and *EP*; *K14-CreER* ^{T2} mice precluded scoring a prostate phenotype.

However, *EP; K8-CreER*^{T2} mice remained well following tamoxifen induction and were aged until 9 months. Despite early and sustained luminal-specific ERG induction (from one week, then monthly through the 9 months observation period) (**Fig. E-G, fig. S10A**), we failed to detect any evidence of invasive cancer (n=8 at 6-9 months). The only histologic changes of note were hyperplasia and PIN in nearly all mice (**Fig. 3H**), a phenotype also seen after Pten loss alone (*35, 36, 43*). We did note rare foci of ERG+ IM cells in a single mouse after 9 months but without evidence of invasive cancer (**fig. S10B**), raising the possibility of leaky Cre expression or expansion of a rare K5+/K8+ DP cell. Although the *K5-CreER*^{T2} half of this experiment was uninformative for scoring a prostate phenotype, the *K8-CreER*^{T2} results indicate that K8+ luminal cells are unlikely to serve as cells of origin, consistent with the OT transplantation studies.

To avoid the lethal skin phenotype that complicates use of *K5-CreER*^{T2} and *K14-CreER*^{T2} mice and directly test whether ERG-driven prostate cancers can initiate from basal cells *in situ*,

we delivered Cre locally by intra-prostatic injection of adenoviruses expressing Cre downstream of K5- vs K8-specific regulatory sequences (Ad-K5-Cre, Ad-K8-Cre) (**Fig. 3I**). To evaluate the fidelity and robustness of this method of Cre delivery in the prostate, we used *Rosa26-YFP*$^{LSL/LSL}$ reporter mice as a control, as we did with the *K5-CreER*^{T2} and *K8-CreER*^{T2} strains. One week after injection of Ad-K5-Cre, we found all YFP+ cells were K5+ with a flat-shaped basal morphology. In contrast, nearly all YFP+ cells in Ad-K8-Cre injected mice were K8/18+ with a columnar luminal morphology (**Fig. 3J**). Having demonstrated the lineage specificity of the Ad-K5 and Ad-K8-Cre viruses, we performed similar injections in *EP* mice (two injections per mouse into either anterior or dorsal lobes), then followed each cohort for disease onset. In Ad-K5-Cre injected mice, we observed clear expansion of ERG+ cells at 12 out of 13 injection sites, with histologic evidence of invasive adenocarcinoma in half (6 of 12) (**Fig. 3K-L, fig. S11A**). Despite targeting ERG activation in K5+ positive basal cells, the ERG+ invasive cells had luminal morphology and expressed K8, K18 and Trop2, together with ERG+ IM (K5+, Ar+) cells within the same region (**Fig. 3M, fig. S11B-C**). In contrast, in mice injected with Ad-K8-Cre virus, we did not observe consistent expansion of ERG+ cells (smaller foci were seen in 2 of 13 mice) and there was no evidence of invasion (**Fig. 3K-M, fig. S11**).

A subset of normal basal cells express luminal genes

Having now provided multiple lines of evidence that K5+ basal cells are a preferred cell of origin for ERG-driven prostate cancers in mice, we considered the fact that ERG expression in ERG-translocated human prostate cancers is typically under the transcriptional control of *TMPRSS2*, which is most highly expressed in luminal cells in the mouse and human prostate. To address this potential inconsistency with a basal cell of origin model, we re-examined mouse and human normal prostate scRNA-seq datasets and noted small numbers of cells within basal clusters express TMPRSS2 (in addition to the uniform expression of TMPRSS2 in luminal subpopulations). Of note, other canonical androgen regulated luminal genes such as *NKX3-1*, *NDRG1*, *FKBP5*, *KLK3* and *Pbsn* were also expressed in a subset of basal cells in normal human and mouse prostate (**Fig. 4A, fig. S12**). To generate functional evidence for this subpopulation of basal cells, we crossed *Nkx3-1*-CreER^{T2} mice into the *Rosa26-YFP*$^{LSL/LSL}$ background. As expected (*38*), the largest population of YFP+ cells were luminal (K8/18+) but we saw clear evidence that a subset of basal cells (K5+, p63+) express YFP. Importantly these YFP+ cells

display the flat morphology typical of basal cells yet are adjacent to luminal cells at the basement membrane (**Fig. 4B-C**). We therefore refer to these as basal-luminal hybrid cells.

ERG drives rapid expansion of basal-luminal hybrid cells

The *Nkx3-1*-CreER^{T2} driver has been used extensively to generate prostate cancer models across a range of genetic contexts (*Pten*$^{flox/flox}$; *Pten*$^{flox/flox}$,*Trp53*$^{flox/flox}$; *Pten*$^{flox/flox}$, *Kras*$^{LSL\text{-}G12D}$; *Pten*$^{flox/flox}$,*Rosa26-ERG*$^{LSL/LSL}$), often implicating luminal cells as cell of origin for these cancers (*36, 46-48*). Having now documented the Nxk3.1 expression in a subpopulation of basal cells (basal-luminal hybrid cells), we crossed *Nkx3-1-CreER*^{T2} mice into the *EP* background to explore the relative expansion of these cells versus the more abundant population of Nkx3.1+ luminal cells after ERG induction. Consistent with the YFP data, a small fraction (11%) of ERG+ cells at one week were basal (K5+, K8/18-) whereas ~85% were luminal. However, a new population of Ki67+, K5+, K8/18+ IM cells emerged one month after ERG induction (**Fig. 4D-F**). Initially (one week after induction), these K5+, ERG+ cells were individual, flat-shaped basal cells adjacent to the basement membrane but one month later appeared as clusters with a mix of K5+, K8/18- single positive and K5+, K8/18+ DP cells (**Fig. 4G**). After 9 months, all *Nkx3-1-CreER*^{T2}*;EP* mice developed invasive adenocarcinomas. We observed IM populations (K5+, Ar+, morphologically luminal) in 12 of 13 lesions examined, as seen in *EPC* mice (**Fig. 4H-I, fig. S13A**). Because we could find no clear evidence of this early expansion of ERG+ cells using the *K8-CreER*^{T2} driver (**Fig. 4D, fig. S13B-C**), we conclude that ERG+ cancers preferentially initiate in Nkx3.1+ basal-luminal hybrid cells.

Evidence of ERG+ basal cells in human prostate cancer

To explore whether the ERG+ basal-luminal hybrid cells seen in these mouse models are also detectable in ERG+ human prostate cancers, we annotated ERG and TMPRSS2 expression in a published scRNA-seq dataset of localized PCa from 11 patients (*49*). As expected, large clusters of ERG+ luminal cancer cells are clearly seen in the ERG+ patients, but we also observed a small number of ERG+ cells in the corresponding basal clusters of these same patients (5 patients) (**Fig. 4J, fig. S14**). These findings from scRNA-seq data are consistent with prior evidence of TMPRSS2-ERG expression in basal cells as a bacterial artificial chromosome

transgene in mice and as primary cultures from ERG-positive human tumors (*50, 51*). Taken together with the mouse findings, this evidence for TMPRSS2-expressing basal-luminal hybrid cells in human raises the possibility that ERG-positive prostate cancers initiate in these cells, which rapidly transition to the fully differentiated luminal cells that account for the overwhelming number of tumor cells evident at clinical diagnosis (**Fig. 4K**).

While our identification of basal-luminal hybrid cells may help resolve prior confusion regarding the preferred cell of origin of ERG-positive prostate cancer, the question of whether ERG-negative cancers also arise from these cells remains to be answered. It is worth noting that several recent human transcriptomic datasets have revealed subsets of human prostate cancers with basal or basal-like expression signatures, often with clinically important prognostic implications, that have not been previously evident by immunohistochemistry (*52-54*). Moving beyond prostate cancer, the notion that cancers can initiate in a cell with a different lineage identity from the resultant tumor was first established in chronic myeloid leukemia, where the BCR-ABL translocation initiates in a hematopoietic stem cell but the clinical manifestation of the disease is elevated levels of terminally differentiated myeloid cells (neutrophils) (*55, 56*). Differences in the lineage identity of initiating cells and resultant tumor cells may also be relevant in other epithelial malignancies such as breast cancer where basal-to-luminal or luminal-to-basal lineage shifts have been reported with different oncogenic drivers (Brca1 loss; Pik3ca mutation) (*57-61*). Whether these lineage transitions are restricted to a subpopulation of basal or luminal cells remains to be determined. An additional consequence of this work is the insight to focus future investigations of ERG biology on the subpopulation of ERG+ cells responsible for disease initiation, specifically basal-luminal hybrid and ERG+ IM cells rather than ERG+ luminal cells.

References

1. S. A. Tomlins *et al.*, Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science* **310**, 644-648 (2005).
2. R. L. Siegel, K. D. Miller, N. S. Wagle, A. Jemal, Cancer statistics, 2023. *CA Cancer J Clin* **73**, 17-48 (2023).
3. N. Cerveira *et al.*, TMPRSS2-ERG gene fusion causing ERG overexpression precedes chromosome copy number changes in prostate carcinomas and paired HGPIN lesions. *Neoplasia* **8**, 826-832 (2006).

4. C. Gerhauser *et al.*, Molecular Evolution of Early-Onset Prostate Cancer Identifies Molecular Risk Markers and Clinical Trajectories. *Cancer Cell* **34**, 996-1011 e1018 (2018).
5. Y. Chen *et al.*, ETS factors reprogram the androgen receptor cistrome and prime prostate tumorigenesis in response to PTEN loss. *Nat Med* **19**, 1023-1029 (2013).
6. B. S. Carver *et al.*, Aberrant ERG expression cooperates with loss of PTEN to promote cancer progression in the prostate. *Nat Genet* **41**, 619-624 (2009).
7. J. C. King *et al.*, Cooperativity of TMPRSS2-ERG with PI3-kinase pathway activation in prostate oncogenesis. *Nat Genet* **41**, 524-526 (2009).
8. S. Perner *et al.*, TMPRSS2-ERG fusion prostate cancer: an early molecular event associated with invasion. *Am J Surg Pathol* **31**, 882-888 (2007).
9. E. Shrestha *et al.*, Oncogenic gene fusions in nonneoplastic precursors as evidence that bacterial infection can initiate prostate cancer. *Proc Natl Acad Sci U S A* **118**, (2021).
10. J. M. Mosquera *et al.*, Characterization of TMPRSS2-ERG fusion high-grade prostatic intraepithelial neoplasia and potential clinical implications. *Clin Cancer Res* **14**, 3380-3385 (2008).
11. R. Bose *et al.*, ERF mutations reveal a balance of ETS factors controlling prostate oncogenesis. *Nature* **546**, 671-675 (2017).
12. Z. Mounir *et al.*, TMPRSS2:ERG blocks neuroendocrine and luminal cell differentiation to maintain prostate cancer proliferation. *Oncogene* **34**, 3815-3825 (2015).
13. S. A. Tomlins *et al.*, Role of the TMPRSS2-ERG gene fusion in prostate cancer. *Neoplasia* **10**, 177-188 (2008).
14. J. Yu *et al.*, An integrated network of androgen receptor, polycomb, and TMPRSS2-ERG gene fusions in prostate cancer progression. *Cancer Cell* **17**, 443-454 (2010).
15. P. C. Hollenhorst *et al.*, Oncogenic ETS proteins mimic activated RAS/MAPK signaling in prostate cells. *Genes Dev* **25**, 2147-2157 (2011).
16. L. T. Nguyen *et al.*, ERG Activates the YAP1 Transcriptional Program and Induces the Development of Age-Related Prostate Tumors. *Cancer Cell* **27**, 797-808 (2015).
17. G. J. Sandoval *et al.*, Binding of TMPRSS2-ERG to BAF Chromatin Remodeling Complexes Mediates Prostate Oncogenesis. *Mol Cell* **71**, 554-566 e557 (2018).
18. J. C. Brenner *et al.*, Mechanistic rationale for inhibition of poly(ADP-ribose) polymerase in ETS gene fusion-positive prostate cancer. *Cancer Cell* **19**, 664-678 (2011).
19. N. Cancer Genome Atlas Research, The Molecular Taxonomy of Primary Prostate Cancer. *Cell* **163**, 1011-1025 (2015).
20. F. Li *et al.*, ERG orchestrates chromatin interactions to drive prostate cell fate reprogramming. *The Journal of clinical investigation* **130**, 5924-5941 (2020).
21. Z. Chen *et al.*, Crucial role of p53-dependent cellular senescence in suppression of Pten-deficient tumorigenesis. *Nature* **436**, 725-730 (2005).
22. W. R. Karthaus *et al.*, Regenerative potential of prostate luminal cells revealed by single-cell analysis. *Science* **368**, 497-505 (2020).
23. L. Crowley *et al.*, A single-cell atlas of the mouse and human prostate reveals heterogeneity and conservation of epithelial progenitors. *Elife* **9**, (2020).
24. W. Guo *et al.*, Single-cell transcriptomics identifies a distinct luminal progenitor cell type in distal prostate invagination tips. *Nat Genet* **52**, 908-918 (2020).
25. A. A. Germanos *et al.*, Defining cellular population dynamics at single-cell resolution during prostate cancer progression. *Elife* **11**, (2022).
26. A. Centonze *et al.*, Heterotypic cell-cell communication regulates glandular stem cell multipotency. *Nature* **584**, 608-613 (2020).

27. O. J. Kwon, L. Zhang, M. M. Ittmann, L. Xin, Prostatic inflammation enhances basal-to-luminal differentiation and accelerates initiation of prostate cancer with a basal cell origin. *Proc Natl Acad Sci U S A* **111**, E592-600 (2014).
28. M. Ousset *et al.*, Multipotent and unipotent progenitors contribute to prostate postnatal development. *Nat Cell Biol* **14**, 1131-1138 (2012).
29. R. Toivanen, A. Mohan, M. M. Shen, Basal Progenitors Contribute to Repair of the Prostate Epithelium Following Induced Luminal Anoikis. *Stem Cell Reports* **6**, 660-667 (2016).
30. A. Wuidart *et al.*, Quantitative lineage tracing strategies to resolve multipotency in tissue-specific stem cells. *Genes Dev* **30**, 1261-1277 (2016).
31. A. S. Goldstein, J. Huang, C. Guo, I. P. Garraway, O. N. Witte, Identification of a cell of origin for human prostate cancer. *Science* **329**, 568-571 (2010).
32. T. Stoyanova *et al.*, Prostate cancer originating in basal cells progresses to adenocarcinoma propagated by luminal-like cells. *Proc Natl Acad Sci U S A* **110**, 20111-20116 (2013).
33. D. A. Lawson *et al.*, Basal epithelial stem cells are efficient targets for prostate cancer initiation. *Proc Natl Acad Sci U S A* **107**, 2610-2615 (2010).
34. J. W. Park *et al.*, Prostate epithelial cell of origin determines cancer differentiation state in an organoid transformation assay. *Proc Natl Acad Sci U S A* **113**, 4482-4487 (2016).
35. N. Choi, B. Zhang, L. Zhang, M. Ittmann, L. Xin, Adult murine prostate basal and luminal cells are self-sustained lineages that can both serve as targets for prostate cancer initiation. *Cancer Cell* **21**, 253-265 (2012).
36. Z. A. Wang *et al.*, Lineage analysis of basal epithelial cells reveals their unexpected plasticity and supports a cell-of-origin model for prostate cancer heterogeneity. *Nat Cell Biol* **15**, 274-283 (2013).
37. Z. A. Wang, R. Toivanen, S. K. Bergren, P. Chambon, M. M. Shen, Luminal cells are favored as the cell of origin for prostate cancer. *Cell Rep* **8**, 1339-1346 (2014).
38. X. Wang *et al.*, A luminal epithelial stem cell that is a cell of origin for prostate cancer. *Nature* **461**, 495-500 (2009).
39. Y. A. Yoo *et al.*, Bmi1 marks distinct castration-resistant luminal progenitor cells competent for prostate regeneration and tumour initiation. *Nat Commun* **7**, 12943 (2016).
40. W. Feng *et al.*, Rapid interrogation of cancer cell of origin through CRISPR editing. *Proc Natl Acad Sci U S A* **118**, (2021).
41. A. Van Keymeulen *et al.*, Distinct stem cells contribute to mammary gland development and maintenance. *Nature* **479**, 189-193 (2011).
42. J. R. Rock *et al.*, Basal cells as stem cells of the mouse trachea and human airway epithelium. *Proc Natl Acad Sci U S A* **106**, 12771-12775 (2009).
43. T. L. Lu *et al.*, Conditionally ablated Pten in prostate basal cells promotes basal-to-luminal differentiation and causes invasive prostate cancer in mice. *Am J Pathol* **182**, 975-991 (2013).
44. L. Zhang *et al.*, Targeting CreER(T2) expression to keratin 8-expressing murine simple epithelia using bacterial artificial chromosome transgenesis. *Transgenic Res* **21**, 1117-1123 (2012).
45. C. Blanpain, Tracing the cellular origin of cancer. *Nat Cell Biol* **15**, 126-134 (2013).
46. M. Zou *et al.*, Transdifferentiation as a Mechanism of Treatment Resistance in a Mouse Model of Castration-Resistant Prostate Cancer. *Cancer Discov* **7**, 736-749 (2017).

47. A. Aytes *et al.*, ETV4 promotes metastasis in response to activation of PI3-kinase and Ras signaling in a mouse model of advanced prostate cancer. *Proc Natl Acad Sci U S A* **110**, E3506-3515 (2013).
48. A. Vasciaveo *et al.*, OncoLoop: A Network-Based Precision Cancer Medicine Framework. *Cancer Discov* **13**, 386-409 (2023).
49. H. Song *et al.*, Single-cell analysis of human primary prostate cancer reveals the heterogeneity of tumor-associated epithelial cell states. *Nat Commun* **13**, 141 (2022).
50. E. S. Polson *et al.*, Monoallelic expression of TMPRSS2/ERG in prostate cancer stem cells. *Nat Commun* **4**, 1623 (2013).
51. O. M. Casey *et al.*, TMPRSS2- driven ERG expression in vivo increases self-renewal and maintains expression in a castration resistant subpopulation. *PLoS One* **7**, e41668 (2012).
52. S. Chen *et al.*, Single-cell analysis reveals transcriptomic remodellings in distinct cell types that contribute to human prostate cancer progression. *Nat Cell Biol* **23**, 87-98 (2021).
53. S. G. Zhao *et al.*, Associations of Luminal and Basal Subtyping of Prostate Cancer With Prognosis and Response to Androgen Deprivation Therapy. *JAMA Oncol* **3**, 1663-1672 (2017).
54. S. You *et al.*, Integrated Classification of Prostate Cancer Reveals a Novel Luminal Subtype with Poor Outcome. *Cancer Res* **76**, 4948-4958 (2016).
55. D. Vetrie, G. V. Helgason, M. Copland, The leukaemia stem cell: similarities, differences and clinical prospects in CML and AML. *Nat Rev Cancer* **20**, 158-173 (2020).
56. M. W. Deininger, J. M. Goldman, J. V. Melo, The molecular biology of chronic myeloid leukemia. *Blood* **96**, 3343-3356 (2000).
57. A. Van Keymeulen *et al.*, Reactivation of multipotency by oncogenic PIK3CA induces breast tumour heterogeneity. *Nature* **525**, 119-123 (2015).
58. S. Koren *et al.*, PIK3CA(H1047R) induces multipotency and multi-lineage mammary tumours. *Nature* **525**, 114-118 (2015).
59. G. Molyneux *et al.*, BRCA1 basal-like breast cancers originate from luminal epithelial progenitors and not from basal stem cells. *Cell Stem Cell* **7**, 403-417 (2010).
60. E. Lim *et al.*, Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers. *Nat Med* **15**, 907-913 (2009).
61. J. E. Visvader, Cells of origin in cancer. *Nature* **469**, 314-322 (2011).

Acknowledgments: We thank the Sawyers lab for valuable critiques and discussions. We appreciate the feedback from Drs. Michael M. Shen and Cory Abate-Shen, and advice from Cory Abate-Shen lab for setting up the intraprostatic adenoviral injection assay. We are grateful to Rona Lester for help in mouse colony management, the Molecular Cytology Core Facility from MSKCC for help with microscopy, and the Flow Cytometry Core Facility from MSKCC for help with FACS experiments.

Funding: This work is supported by the following funding sources.

Howard Hughes Medical Institute (HHMI) (CLS)

National Institute of Health grants CA193837 (CLS.), CA092629 (CLS), CA224079 (CLS), CA155169 (CLS), 1K99CA276888-01 (WF)

Department of Defense grant W81XWH-19-1-0323 (WF)

Prostate Cancer Foundation 20YOUN22 (WF)

Author contributions:

Conceptualization: WF, CLS

Methodology: WF, EL, HZ, HL, CSL

Investigation: WF, EL, NS, HZ, YSL, AG, WK, NF, ER

Formal analysis: WF, EL

Resources: YC, BSC

Project administration: IM, RC, ES

Funding acquisition: WF, CLS

Supervision: CSL, CLS

Writing – original draft: WF, CLS

Writing – review & editing: WF, CLS

Competing interests: Dr. Sawyers serves on the Board of Directors of Novartis, is a co-founder of ORIC Pharmaceuticals and co-inventor of enzalutamide and apalutamide. He is a science advisor to Arsenal, Beigene, Blueprint, Column Group, Foghorn, Housey Pharma, Nextech, KSQ and PMV.

Data and materials availability: All data are available in the main text or the supplementary materials.

Supplementary Materials

Materials and Methods

Figs. S1 to S14

Tables S1 to S2

References

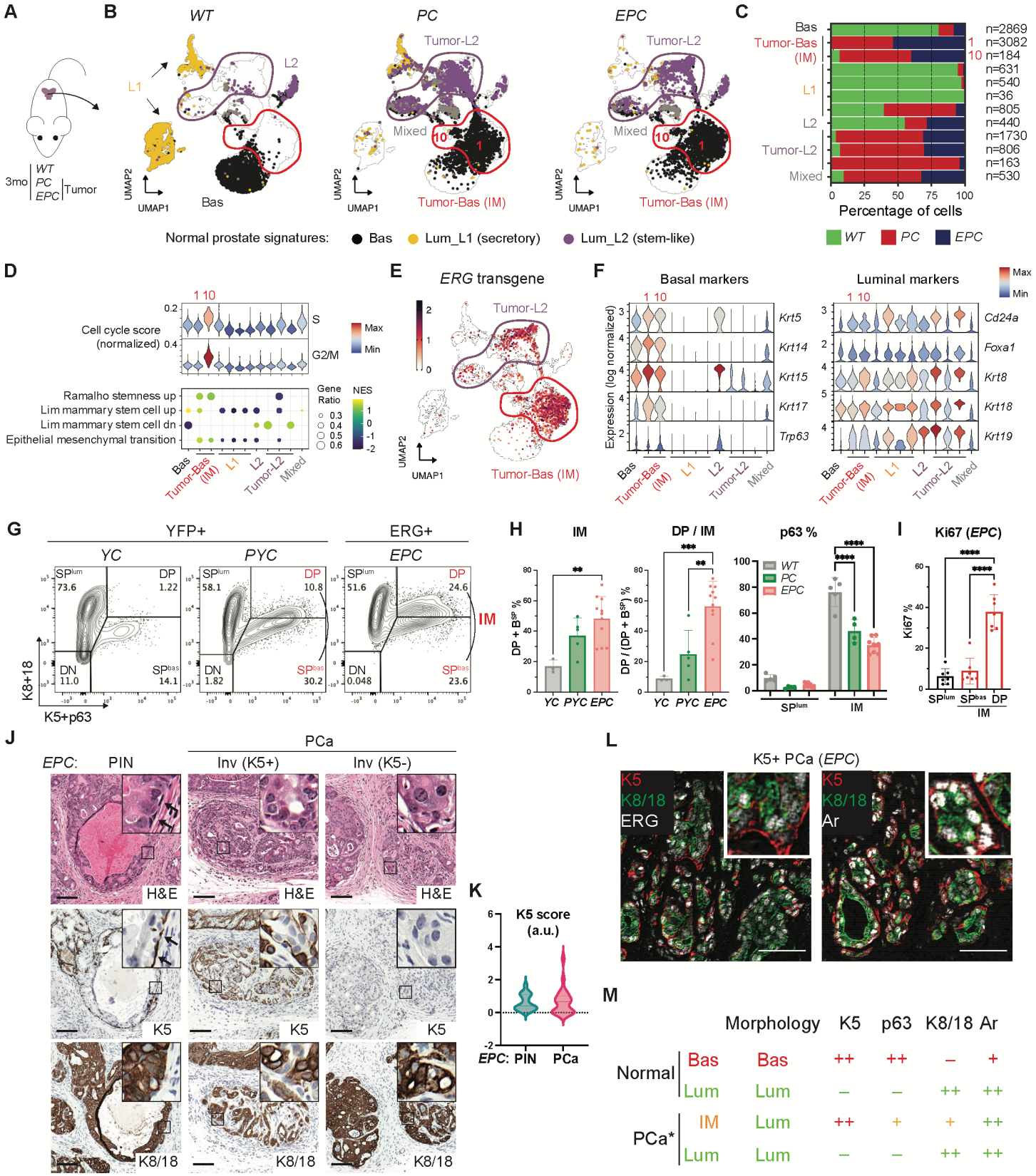

Fig. 1. Presence of a highly proliferative intermediate state in ERG-driven prostate cancer.

(**A**) Experimental design including *WT* (1 mouse), *PC* (2 mice, *Pten*$^{flox/flox}$*;Pb-Cre4*) and *EPC* (2 mice, *Rosa26-ERG*$^{LSL/LSL}$*; Pten*$^{flox/flox}$*;Pb-Cre4*). (**B-C**) UMAP (**B**) and barplot (**C**) showing epithelial clusters assigned to each genotype. Numbers of cells in each cluster were included (N = 11,816 epithelial cells in total). Cluster annotation was based on normal mouse prostate epithelial signatures (*22*). L2 and Bas clusters specific to tumor samples (*PC* and *EPC*) are highlighted in circles. The IDs of the two Tumor-Bas clusters are shown, which together were subsequently defined as intermediate cells (IM). A complete list of all cluster IDs are shown in **fig. S1E**. A cluster containing a mix of different cell types is defined as mixed. Bas, basal; Lum, luminal (**D**) Cell cycle scores (top) and gene set enrichment analysis (bottom) across all epithelial clusters. The two Tumor-Bas (IM) clusters are numbered according to the UMAP in **B**. Pathways enriched with a FDR-adjusted p-value <0.05 are shown. (**E**) UMAP plotting the *ERG* transgene expression. (**F**) Violin plots comparing canonical basal and luminal marker expression across all epithelial clusters. The two Tumor-Bas (IM) clusters are numbered according to the UMAP in **B**. (**G**) Intracellular flow cytometry assessing basal (K5, p63) and luminal (K8, K18) marker expression. Recombined cells labeled by YFP or ERG were analyzed for *YC* (*Rosa26-YFP*$^{LSL/LSL}$*;Pb*-Cre4)/*PYC* (*Pten*$^{flox/flox}$*; Rosa26-YFP*$^{LSL/LSL}$*;Pb-Cre4*) and *EPC* mice, respectively. The combined SP^{bas} and DP fractions in tumors are referred to as IM cells. SP, single positive; DP, double positive. (**H**) Flow cytometry quantification highlighting an expansion of the IM population in tumors and a further luminal shift in IM cells with ERG expression. (**I**) Flow cytometry quantification of Ki67+ cells confirming a high proliferation in the DP fraction of IM cells. (**J**) Haematoxylin–eosin (H&E) prostate histology and immunohistochemistry (IHC) of 3-month *EPC* mice highlighting a luminal morphology of both K5+ and K5- cells from invasive adenocarcinomas, in contrast to the K5+ basal cells that encapsulate the precursor PIN lesions (arrow); inset shows high-power view. PIN, prostatic intraepithelial neoplasia; inv, invasive adenocarcinoma; PCa, prostate cancer. Scale bars, 100 µm. (**K**) Quantification of K5 IHC signal in PIN and invasive adenocarcinomas of *EPC* prostates at 2~3-month age. (**L**) Immunofluorescence (IF) staining showing expression of ERG (left) and Ar (right) in both K5- and K5+ cells from invasive adenocarcinomas; inset shows high-power view. Scale bars, 50 µm. (**M**) Summary of cell morphology and protein marker expression across different cell types based on flow cytometry and immunostaining in **G-L**. Colors are assigned based on normal basal and luminal cells, which highlight the basal-luminal hybrid features of IM cells in PCa. Asterisk, basal cells are absent in PCa, which displays IM cells instead. Data represent mean ± s.d.; n ≥ 3.; **p < 0.01; ***p<0.001; ****p < 0.0001; one-way ANOVA with Tukey posttest.

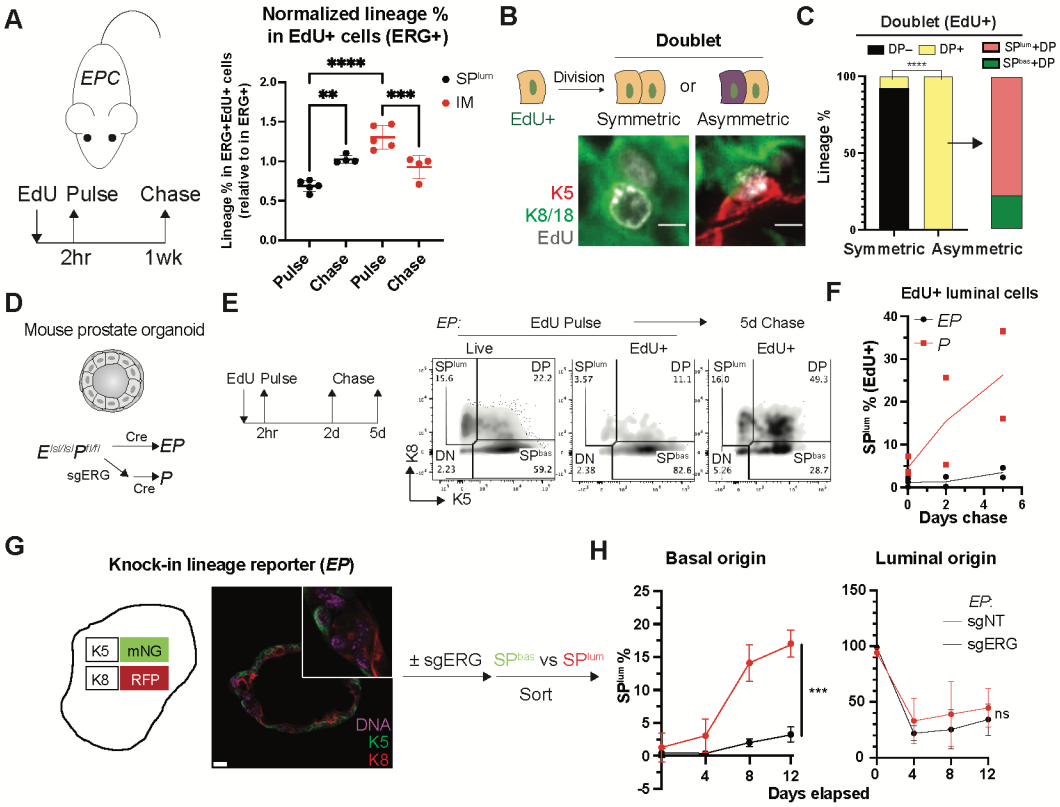

Fig. 2. Intermediate and basal cells transit to a luminal fate in the presence of ERG.

(**A**) EdU pulse-chase assay on *EPC* mice at 2-month age (schematic on the left). Lineage representation in the EdU+ population of ERG+ cells (normalized to all ERG+ cells) on the right shows an increase of SP^{lum} cells and decrease of IM after the chase. (**B**) Schematic (top) and representative immunofluorescence images (bottom) exemplifying symmetric and asymmetric divisions of EdU-labelled cells. Scale bars, 5 µm. (**C**) The cell type composition of EdU+ doublets highlight that DP cells are present in all asymmetric divisions and preferentially associate with a SP^{lum} daughter cell. (**D**) Schematic showing *in vitro* generation of an isogenic pair of *EP* and *P* organoids. (**E,F**) EdU pulse-chase assay in organoids followed by flow cytometry assessment highlight an ERG-dependent shift of the EdU-labeled cells towards a luminal (SP^{lum}) fate. Representative data (**E**) and quantification of the luminal fraction in EdU-labeled populations (**F**) are shown. (**G**) (Left) Schematic showing *EP* organoids with a dual lineage reporter knocked in to form C-terminus fusions with endogenous K5 and K8. (Middle) Live cell imaging highlighting the expected spatial expression pattern of the reporter signals, which enables live sort of basal and luminal cells (schematic on the right). mNG, mNeonGreen; RFP, TagRFP. Scale bar, 20 µm. (**H**) The basal- and luminal-derived *EP* reporter organoids from **G** were analyzed by flow cytometry based on the reporter signals along the time course post sort. An ERG-dependent luminal fate transition from basal cells from luminal cells was observed. Data represent mean ± s.d.; $n \geq 3$ (except $n = 2$ in **F**).; ns, not significant, $*p < 0.05$; $**p < 0.01$; $***p<0.001$; $****p < 0.0001$; one-way ANOVA with Tukey posttest (**A**); Fisher's exact test (**C**); two-way ANOVA (**H**).

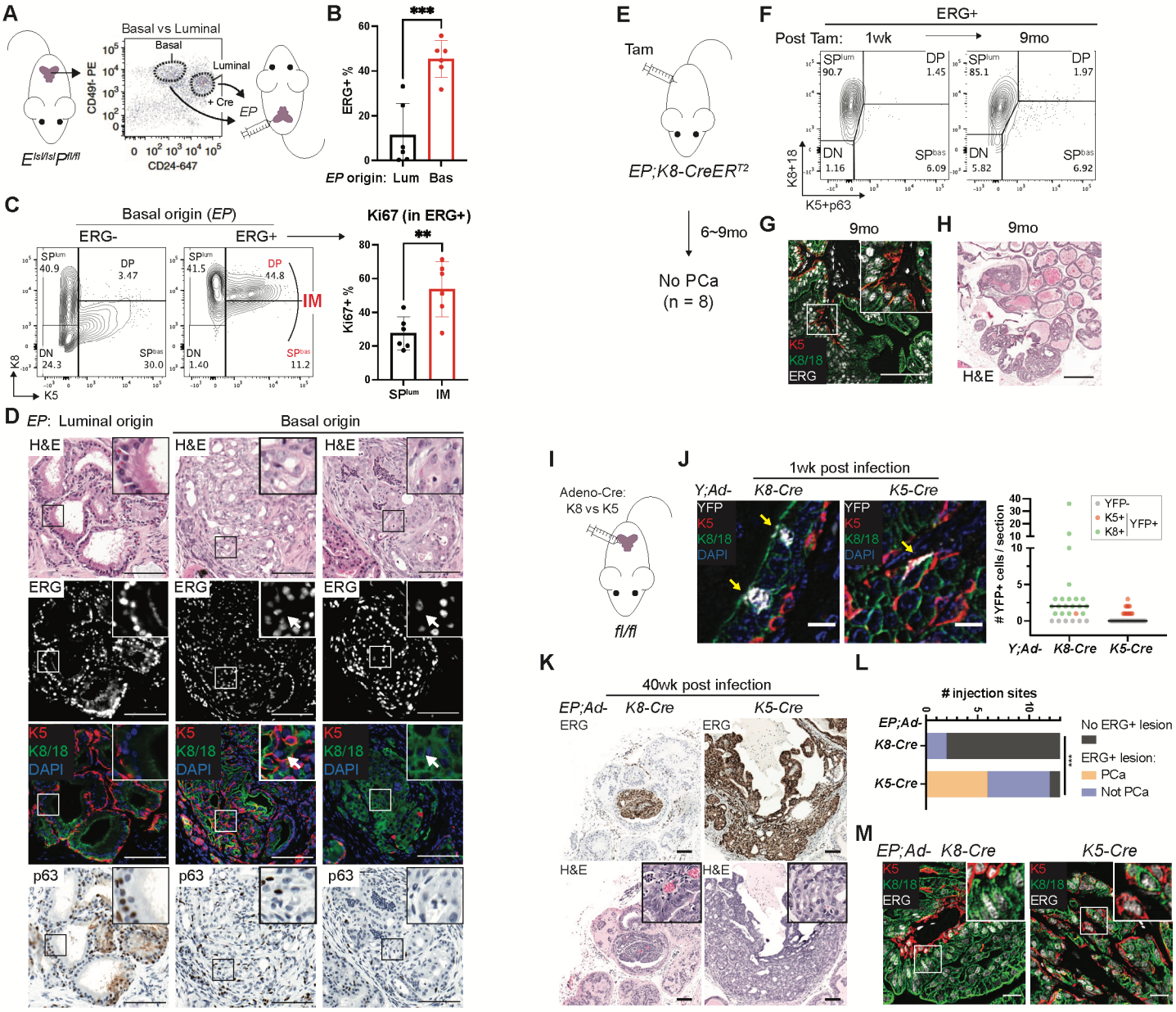

Fig. 3. Basal cells are preferred to form ERG-driven prostate cancer with luminal features.

(**A**) An orthotopic transplantation approach to compare basal- and luminal-derived *EP* orthografts using freshly isolated and recombined prostate cells. (**B**) Flow cytometry quantification of ERG+ cells in grafts harvested at the 5-month endpoint showing a higher graft burden from the basal-derived orthografts. (**C**) Intracellular flow cytometry highlighting the presence of IM population (left) which were more proliferative than luminal (SP^{lum}) cells by Ki67 (right) in ERG+ grafted cells. Mice were harvested at the 5-month endpoint. (**D**) Prostate histological analysis highlighting a histological recapitulation of invasive adenocarcinomas and the presence of luminal cells (K8/18+K5-p63-, arrow) from 5-month basal-originated *EP* grafts; inset shows high-power view. Scale bars, 100 µm. (**E**) A lineage tracing approach to activate *EP* by tamoxifen using a luminal-specific *K8-CreER*^{T2} driver. No PCa (defined as invasive adenocarcinoma) was observed after 9 months. (**F**) Flow cytometry highlighting luminal-specific ERG expression from onset (1 week) and maintained through 9 months of tracing in *EP; K8-*

CreER^{T2} mice. (**G-H**) Prostate histological analysis showing luminal identity (K8/18+K5-) of ERG+ cells (**G**) and lack of PCa development (**H**) in *EP; K8-CreER^{T2}* mice up to 9 months of tracing; inset shows high-power view. Scale bars, 100 µm (**G**), 500 µm (**H**). (**I**) An intra-prostatic adenoviral (Ad) injection approach to compare *EP* activation by Ad-K8- vs Ad-K5-Cre. (**J**) IF images (left) and quantification (right) assessing the cell type specificity and infection efficiency of the indicated Ad-Cre at the 1-week time point. Scale bars, 10 µm. (**K-L**) Histology (**K**) and characterization (**L**) of *EP* prostates harvested at 40 weeks post infection; inset, a high-power view highlighting an overall luminal morphology of the invasive cancer cells from *EP;Ad-K5-Cre* mice. PCa is defined as invasive adenocarcinoma. Scale bars, 100 µm. (**M**) IF of *EP* prostates harvested at 40 weeks post infection highlighting the presence of luminal cells (K8/18+K5-) in invasive adenocarcinomas developed from *EP;Ad-K5-Cre* mice; inset shows high-power view. Scale bars, 20 µm. Data represent mean ± s.d.; $n \geq 3$.; ***$p<0.001$; unpaired two-tailed t-test (**B**); Chi-square test (**L**).

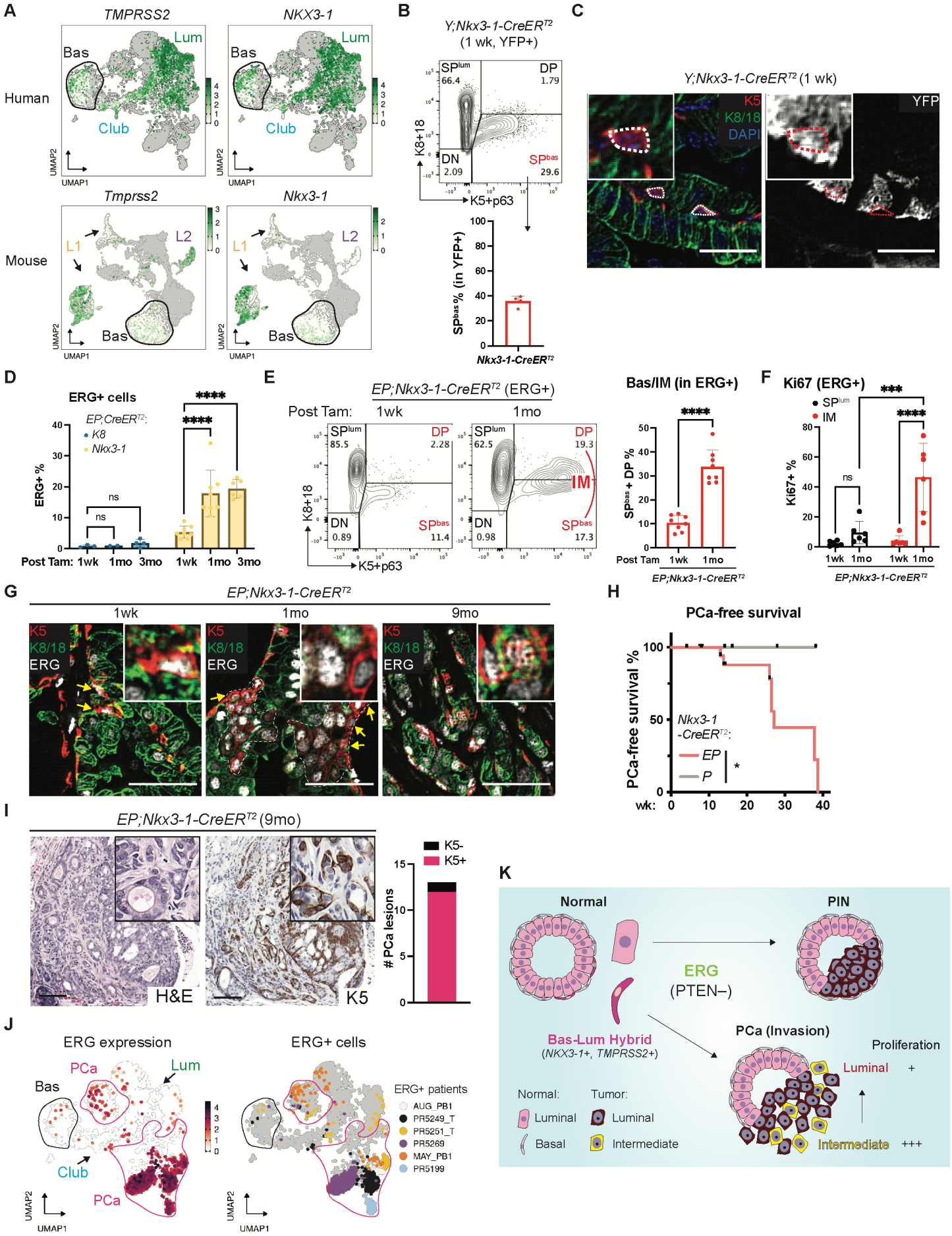
A
TMPRSS2
NKX3-1
Human
Bas
Lum
Club
UMAP2
UMAP1
Tmprss2
Nkx3-1
Mouse
L1
L2
B
Y;Nkx3-1-CreER^{T2}
(1 wk, YFP+)
SP^{lum}
66.4
DP
1.79
DN
2.09
SP^{bas}
29.6
K8+18
K5+p63
SP^{bas} % (in YFP+)
Nkx3-1-CreER^{T2}
C
Y;Nkx3-1-CreER^{T2} (1 wk)
K5
K8/18
DAPI
YFP
D
ERG+ cells
EP;CreER^{T2}:
K8
Nkx3-1
ERG+ %
ns
****
Post Tam: 1wk 1mo 3mo 1wk 1mo 3mo
E
EP;Nkx3-1-CreER^{T2} (ERG+)
Post Tam:
1wk
1mo
SP^{lum}
85.5
DP
2.28
DN
0.89
SP^{bas}
11.4
SP^{lum}
62.5
DP
19.3
IM
DN
0.98
SP^{bas}
17.3
Bas/IM (in ERG+)
SP^{bas} + DP %
Post Tam 1wk 1mo
EP;Nkx3-1-CreER^{T2}
F
Ki67 (ERG+)
SP^{lum}
IM
***
****
Ki67+ %
ns
1wk 1mo 1wk 1mo
EP;Nkx3-1-CreER^{T2}
G
EP;Nkx3-1-CreER^{T2}
1wk
1mo
9mo
K5
K8/18
ERG
H
PCa-free survival
PCa-free survival %
Nkx3-1
-CreER^{T2}:
EP
P
*
wk: 0 10 20 30 40
I
EP;Nkx3-1-CreER^{T2} (9mo)
H&E
K5
K5-
K5+
PCa lesions
J
ERG expression
ERG+ cells
PCa
Lum
Bas
Club
ERG+ patients
AUG_PB1
PR5249_T
PR5251_T
PR5269
MAY_PB1
PR5199
K
Normal
PIN
ERG
(PTEN–)
Bas-Lum Hybrid
(NKX3-1+, TMPRSS2+)
PCa (Invasion)
Proliferation
Luminal +
Intermediate +++
Normal:
Tumor:
Luminal
Basal
Luminal
Intermediate

Fig. 4. ERG-driven prostate cancer expands from a basal subset with basal-luminal hybrid features.

(**A**) UMAP of epithelial cells from normal human and mouse prostates highlighting the expression of canonical luminal genes (*TMPRSS2*, *NKX3-1*, and more in **fig. S12B**) in a subset of basal cells. Cell types are annotated based on **fig. S1E** (mouse) and **fig. S12A** (human), with basal clusters highlighted in circles. Cells from normal samples are colored from white to green based on gene expression, with cells from tumor samples in grey in the background. (**B**) (Left) Flow cytometry highlighting the presence of basal cells from the recombined YFP+ population generated by Nkx3-1-CreER^{T2}. *Y;Nkx3-1-CreER^{T2}* mice were treated with tamoxifen at 2 months and harvested 1 week later. (Right) The fraction of basal cells within the YFP+ population are shown (basal proportion may be overrepresented because basal cells survive tissue dissociation better than luminal cells). (**C**) YFP+ cells that co-express K5 from *Y;Nkx3-1-CreER^{T2}* mice show a flat basal morphology at the gland periphery; inset shows high-power view. Scale bars, 25 µm. (**D**) Flow cytometry quantification of ERG+ cells from *EP* mice activated by the indicated CreER^{T2} drivers along the time course post tamoxifen. (**E**) Flow cytometry from the ERG+ population of the *EP;Nkx3-1-CreER^{T2}* mice showing IM expansion by 1 month post tamoxifen. (**F**) Flow cytometry quantification of Ki67+ cells from the ERG+ population of the *EP;Nkx3-1-CreER^{T2}* mice harvested at the indicated time points. (**G**) IF highlighting the ERG+ cells that co-express K5 along the time course post tamoxifen. Dashed line encircles clusters of K5+ cells. Arrows highlight the K5-single positive (K8/18-negative) cells at the gland periphery. Inset shows high-power view. Scale bars, 50 µm. (**H**) PCa-free survival analysis. PCa is defined as invasive adenocarcinoma. (**I**) Prostate histology (left) and quantification (right) of K5+ cells in PCas (invasive adenocarcinomas) from *EP;Nkx3-1-CreER^{T2}* mice at 9 months post tamoxifen. High-power view (insets) highlights the overall luminal morphology of invasive cells. Scale bars, 100 µm. (**J**) UMAP of epithelial cells from patients highlighting the presence of ERG+ cells in the basal cluster. ERG+ cells are colored by ERG expression (left) and patient ID (right). ERG+ PCa samples are shown (see methods). Cell types are annotated based on **fig. S12A**, with basal and PCa clusters highlighted in black and pink circles, respectively. (**K**) Model: ERG translocations that occur in basal cells with luminal features (Bas-Lum Hybrid cells) enter a highly proliferative intermediate state that, in turn triggers invasive cancer with luminal differentiation. ERG translocations that originate in luminal cells may develop PIN lesions but fail to progress to invasive cancer. Data represent mean ± s.d.; n ≥ 3.; ns, not significant; *p<0.05; ***p<0.001; ****p < 0.0001; two-way ANOVA with Tukey posttest (**D**); unpaired two-tailed t-test (**E**); two-way ANOVA with Sidak posttest (**F**); Log-rank test (**H**).