nature cell biology

Lineage analysis of basal epithelial cells reveals their unexpected plasticity and supports a cell-of-origin model for prostate cancer heterogeneity

Zhu A. Wang[1,2], Antonina Mitrofanova[2,3], Sarah K. Bergren[1,2], Cory Abate-Shen[2,4], Robert D. Cardiff[5], Andrea Califano[2,3] and Michael M. Shen[1,2,6]

A key issue in cancer biology is whether oncogenic transformation of different cell types of origin within an adult tissue gives rise to distinct tumour subtypes that differ in their prognosis and/or treatment response. We now show that initiation of prostate tumours in basal or luminal epithelial cells in mouse models results in tumours with distinct molecular signatures that are predictive of human patient outcomes. Furthermore, our analysis of untransformed basal cells reveals an unexpected assay dependence of their stem cell properties in sphere formation and transplantation assays versus genetic lineage tracing during prostate regeneration and adult tissue homeostasis. Although oncogenic transformation of basal cells gives rise to tumours with luminal phenotypes, cross-species bioinformatic analyses indicate that tumours of luminal origin are more aggressive than tumours of basal origin, and identify a molecular signature associated with patient outcome. Our results reveal the inherent plasticity of basal cells, and support a model in which different cells of origin generate distinct molecular subtypes of prostate cancer.

The analysis of tumour cell of origin requires a detailed understanding of tissue cell types and their position in the lineage hierarchy[1]. In particular, stem cells are often considered to be excellent candidate cells of origin for cancer, given their inherent ability to self-renew. In the prostate gland, the three epithelial cell types are luminal cells, which express cytokeratins (CK) 8 and 18, and high levels of androgen receptor, basal cells, which express p63, CK5 and CK14, and rare neuroendocrine cells; in addition, a minor basal subpopulation known as intermediate cells co-express basal and luminal markers[2]. Notably, the adult prostate can undergo cycles of regression and regeneration following androgen ablation and restoration, implying that the prostate epithelium contains stem cells that function to promote regeneration.

So far, prostate stem cell populations have been identified in both the basal and luminal layers[3–7]. In particular, subpopulations of basal cells isolated using cell-surface markers exhibit multipotency and self-renewal in sphere formation as well as tissue reconstitution assays[8–13]. Other work has identified a rare luminal population of castration-resistant Nkx3.1-expressing cells (CARNs) that exhibits stem cell properties in genetic lineage tracing and tissue reconstitution assays[14]. It has been unclear whether these findings are mutually consistent, given the distinct assays for stem cell properties that have been employed.

The cell-of-origin model for intertumour heterogeneity proposes that tumour initiation from distinct cell types in the lineage hierarchy gives rise to tumour subtypes with different prognoses and/or treatment responses[1,15]. Although this model has received considerable support in studies of breast cancer[16], it has not been systematically investigated in prostate cancer. However, several groups have investigated whether luminal cells or basal cells, or both, might serve as cell types of origin for prostate cancer. In particular, lineage-tracing analyses of CARNs have provided evidence that rare luminal cells can act as a cell of origin *in vivo*[14], whereas other studies have shown that lentiviral overexpression of oncogenes in isolated mouse and human basal cells can give rise to tumours with luminal phenotypes in renal grafts, although luminal cells fail to generate tumours under these conditions[13,17]. In addition, a recent study has shown that both luminal and basal cells can serve as cells of origin for prostate cancer, generating tumours that are histologically similar in mouse models[18].

These previous studies have raised the possibility that *ex vivo* cell culture and tissue grafting assays may yield different results from

[1]Departments of Medicine and Genetics and Development, Columbia University College of Physicians and Surgeons, New York, New York 10032, USA. [2]Herbert Irving Comprehensive Cancer Center, Columbia University College of Physicians and Surgeons, New York, New York 10032, USA. [3]Department of Biomedical Informatics and Center for Computational Biology and Bioinformatics, Columbia University College of Physicians and Surgeons, New York, New York 10032, USA. [4]Departments of Urology and Pathology and Cell Biology, Columbia University College of Physicians and Surgeons, New York, New York 10032, USA. [5]Center for Comparative Medicine and Department of Pathology, School of Medicine, University of California, Davis 95616, USA.
[6]Correspondence should be addressed to M.M.S. (e-mail: mshen@columbia.edu)

Received 30 May 2012; accepted 21 January 2013; published online 24 February 2013; DOI: 10.1038/ncb2697

in vivo lineage-tracing analyses. Therefore, we have undertaken a comprehensive analysis of prostate basal cell properties using genetic lineage marking to examine the properties of the identical cell population in multiple assays for stem cell function. Our results show that apparent discrepancies in the published literature can be explained by the considerable plasticity of basal cells in distinct functional assays. Moreover, although both basal and luminal cells can serve as cells of origin for prostate cancer, giving rise to tumours with similar histological phenotypes, our molecular and bioinformatic analysis shows that the luminal origin tumours are more aggressive, and identifies a molecular signature that has predictive value for human patient survival. Thus, our study supports the cell-of-origin model as a basis for distinct prostate cancer subtypes.

RESULTS

Analysis of lineage-marked prostate basal cells in cell culture and grafting assays

To provide a comprehensive analysis, we have performed genetic marking of prostate epithelial basal cells using a *CK5–CreER^{T2}* transgenic line[19] in combination with the *R26R–YFP* reporter allele[20] for isolation of a purified cell population for sphere formation and tissue reconstitution assays and for lineage tracing *in vivo*. In control experiments, tamoxifen induction of hormonally intact *CK5–CreER^{T2} ; R26R–YFP/+* mice resulted in highly specific expression of YFP in 24.5% ($n = 1{,}538/6{,}267$) of CK5$^+$ basal cells in the anterior prostate lobe, whereas no YFP-positivity was observed in non-basal cells ($n = 0/15{,}846$; Fig. 1a); quantification for all experiments is detailed in Supplementary Table S1. We verified that the YFP-marked cells were positive for the basal cell markers p63 and CK14, and were mostly negative for the luminal marker CK18 (Supplementary Fig. S1a,e–n), and were detected at similar frequencies in the dorsolateral (23.2%) and ventral (24.9%) prostate lobes (Supplementary Fig. S1c,d). Furthermore, 1.6% of p63$^+$YFP$^+$ cells also expressed the luminal marker CK18 (Supplementary Fig. S1o–s), indicating that the marked population includes intermediate cells.

Next, we isolated lineage-marked YFP$^+$ cells by flow sorting of dissociated prostate cells from tamoxifen-induced *CK5–CreER^{T2} ; R26R–YFP/+* mice. Approximately 3.2% of total prostate cells were isolated in the YFP$^+$ fraction (Fig. 1b), and greater than 98% of these isolated YFP$^+$ cells were CK5$^+$, p63$^+$ and CK18$^-$ (Supplementary Fig. S2a–l). Furthermore, we compared this YFP$^+$ population to that of Lin$^-$Sca-1$^+$CD49f^{hi} cells, which have been previously characterized as a basal population enriched for stem/progenitor cells[8,12]. Nearly all YFP$^+$cells (98.7%) were contained in the Lin$^-$CD49f$^+$ population, and 8.0% of the YFP$^+$ cells were present in the Lin$^-$Sca-1$^+$CD49f^{hi} fraction (Fig. 1c). Conversely, we found that 24.4% of the Lin$^-$Sca-1$^+$CD49f^{hi} cells from tamoxifen-induced *CK5–CreER^{T2} ; R26R–YFP/+* mice were YFP$^+$ (Fig. 1d,e), similar to the overall percentage of YFP$^+$ basal cells. These results indicate that the YFP$^+$ population includes Lin$^-$Sca-1$^+$CD49f^{hi} cells in a proportional and unbiased manner.

Using unsorted dissociated cells from tamoxifen-induced *CK5–CreER^{T2} ; R26R–YFP/+* prostate tissue in sphere formation assays, we found that approximately 23% of the resulting spheres were YFP$^+$ (Fig. 1f), consistent with the overall percentage of marked basal cells. After flow sorting, 4.7% of the YFP$^+$ cells could form spheres that exhibited expression of basal and luminal markers (Fig. 1g and Supplementary Fig. S2m–r), consistent with previous studies[8,21]. This frequency of sphere formation was similar between the Lin$^-$Sca-1$^+$CD49f^{hi} YFP$^+$ cells and the remaining YFP$^+$ cells (Supplementary Fig. S2s), and was similar in mice at both 2 months and 12 months of age (Fig. 1g).

Tissue reconstitution assays have been used to show that Lin$^-$Sca-1$^+$CD49f$^+$ and Lin$^-$Sca-1$^+$CD49f^{hi}Trop2^{hi} cells readily generate prostate ducts in renal grafts[8,12], but the overall efficiency of basal cells in this assay has not been previously determined. Therefore, we have used our highly purified YFP$^+$ population in prostate reconstitution assays, and have performed a limiting dilution analysis to determine the percentage of basal cells with reconstituting ability. This analysis shows that approximately 3.9% of basal cells contain graft-forming activity (Fig. 1h,i); these grafts exhibit their donor origin by expression of YFP, display prostatic histology and form ducts with luminal, basal and neuroendocrine cells (Fig. 1j–o and Supplementary Fig. S2t). Notably, the similar efficiencies of tissue reconstitution (3.9%) and sphere formation (4.7%) raise the possibility that these two distinct assays may identify the same progenitor population.

Identification of rare bipotential basal cells during prostate regeneration *in vivo*

Using lineage tracing, we explored the properties of basal cells during prostate regeneration *in vivo*, using a protocol similar to one that we have previously employed[14]. We marked prostate basal cells by tamoxifen treatment of hormonally intact *CK5–CreER^{T2} ; R26R–YFP/+* mice, as in Fig. 1a, followed by androgen deprivation to induce prostate regression, and then androgen restoration to promote prostate regeneration (Fig. 2a). After regeneration, we found that most YFP$^+$ cells were basal, but 0.04% of YFP$^+$ cells were luminal ($n = 5/11{,}427$ cells; Fig. 2b–f and Supplementary Table S1). Notably, the percentage of basal cells exhibiting bipotentiality as detected by lineage tracing *in vivo* is significantly lower than the percentage exhibiting stem cell properties in sphere formation or tissue reconstitution assays ($P < 0.0001$; χ^2 test).

We next investigated whether this low detection rate of marked luminal cells might reflect an inherently low proliferation rate of basal cells (Fig. 2g,h and Supplementary Fig. S3). Using both Ki67 immunostaining and BrdU incorporation methods, we found that approximately 8.0% of basal cells proliferated during prostate regeneration (Fig. 2g,h and Supplementary Fig. S3). Furthermore, rare bipotential basal cells were observed at approximately 0.05% frequency using alternative regeneration protocols or in aged mice (Supplementary Fig. S4). Taken together, these results indicate that the frequency of luminal cell generation by dividing basal cells is low but stable during regeneration *in vivo* (Supplementary Table S1).

Given these findings, we then investigated whether lineage-marked luminal progeny of basal cells would accumulate during serial rounds of androgen deprivation and androgen restoration to drive prostate regression and regeneration. We analysed the frequency of YFP$^+$ luminal cells among total YFP$^+$ cells after three and five rounds of serial regeneration, as well as after four rounds in the regressed state (Fig. 2i–o). We observed 0.6% marked luminal cells after three rounds and 3.4% after five rounds (Fig. 2j–l,n,o). Thus, there is a gradual increase in the percentage of luminal cells among

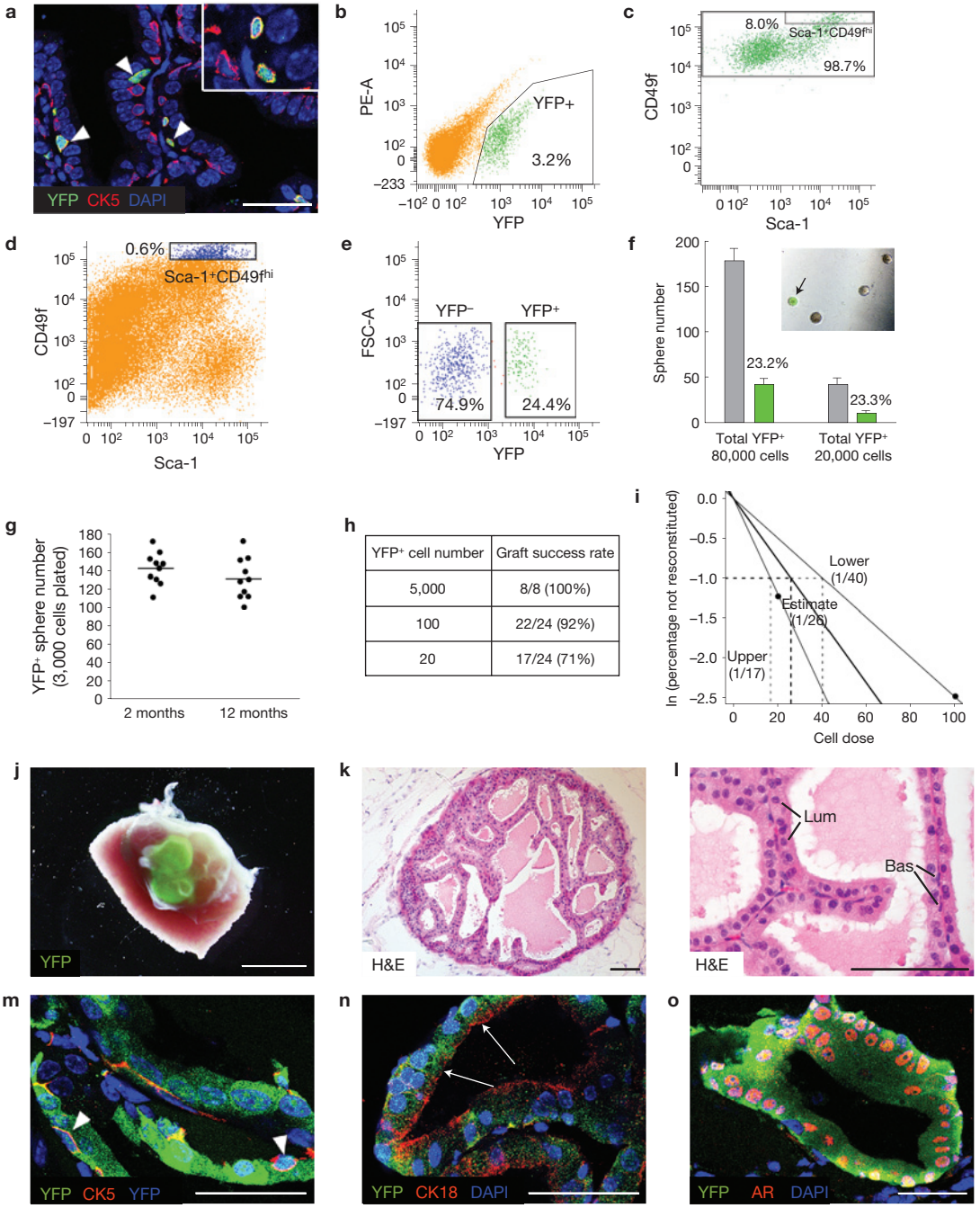

Figure 1 High frequency of prostate basal stem/progenitor cells in sphere formation and tissue reconstitution assays. For all analyses, tamoxifen-induced *CK5–CreER*^{T2} *; R26R–YFP/+* mice were analysed at 14 days after tamoxifen treatment. (**a**) Immunofluorescence staining showing co-localization of YFP with CK5 in basal cells (arrowheads) of the anterior prostate; inset shows high-power view. (**b**) Purification of YFP^{+} basal cells from dissociated prostate tissue by flow cytometry. (**c**) Flow sorting of YFP^{+} cells shows that 98.7% are CD49f^{+}, and 8.0% are Lin^{-}Sca-1^{+}CD49f^{hi} cells. (**d**,**e**) Flow sorting of Lin^{-}Sca-1^{+}CD49f^{hi} cells (**d**, 0.6% of total Lin^{-} cells) shows that 24.4% are YFP^{+} (**e**). (**f**) Quantification of sphere formation from 80,000 or 20,000 dissociated prostate cells, showing the number of total spheres as well as YFP^{+} spheres. Each experiment was performed six times, using two replicates each from 3 independent mice; error bars correspond to standard deviation and show variability between the six samples. The inset shows epifluorescence detection of YFP expression in spheres (arrow). (**g**) Quantification of sphere formation from 3,000 dissociated YFP^{+}cells isolated from *CK5–CreER*^{T2} *; R26R–YFP/+* mice at 2 months or 12 months of age. The frequencies at these two stages are not statistically different by two-sample *t*-test. (**h**) Serial dilution analysis of purified YFP^{+} cells in assays of prostate duct formation in renal grafts. (**i**) Extreme limiting dilution analysis of the data in **h**. (**j**) YFP fluorescence of a renal graft attached to a portion of kidney tissue. (**k**,**l**) Haematoxylin–eosin (H&E) staining of a tissue section from a renal graft generated from purified YFP^{+} basal cells; luminal (Lum) and basal (Bas) cells are indicated (**l**). (**m**–**o**) Analysis of YFP together with CK5 expression in basal cells (arrowheads, **m**), CK18 in luminal cells (arrows, **n**) and AR (**o**) in renal grafts. Scale bars, 50 μm (**a**,**m**–**o**); 100 μm (**k**,**l**); 1 mm (**j**).

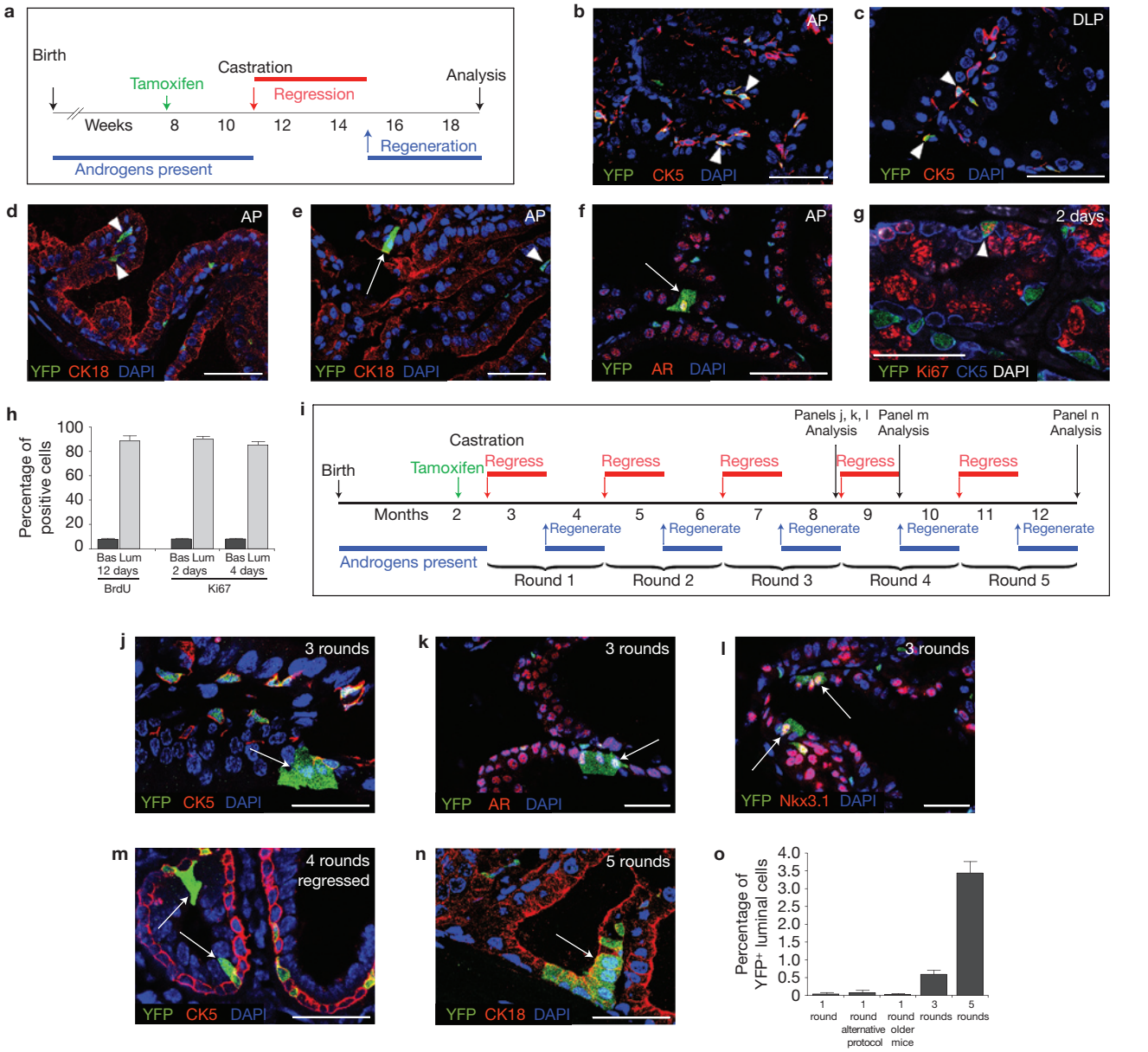

Figure 2 Detection of rare bipotential basal progenitors during prostate regeneration *in vivo*. (**a**) Lineage-tracing strategy during prostate regeneration of *CK5–CreER*^{T2} *; R26R–YFP/+* mice. (**b**,**c**) Co-localization of YFP with CK5-expressing basal cells (arrowheads) in regenerated *CK5–CreER*^{T2} *; R26R–YFP/+* anterior prostate (**b**) and dorsolateral prostate (**c**). (**d**–**f**) Most YFP$^+$ cells (arrowheads) do not express the luminal marker CK18, although rare YFP$^+$ CK18$^+$ (arrow, **e**) and YFP$^+$AR$^+$ (arrow, **f**) luminal cells can be detected after one round of regeneration. (**g**) Ki67 immunostaining 2 days after androgen administration shows that most luminal and some basal cells (arrowhead) undergo proliferation. (**h**) Quantification of cell proliferation assays during regeneration, showing that 7.4% of basal cells ($n = 510$ out of 6,929) and 88.7% of luminal cells ($n = 6{,}708$ out of 7,565) were BrdU$^+$ after 12 days of incorporation; 8.0% of basal cells ($n = 426$ out of 5,326) and 90.1% of luminal cells ($n = 6{,}945$ out of 7,709) were Ki67$^+$ at 2 days of regeneration; and 8.1% of basal cells ($n = 380$ out of 4,700) and 85.1% of luminal cells ($n = 7{,}094$ out of 8,333) were Ki67$^+$ at 4 days of regeneration; 3 animals were analysed for each experiment. See Supplementary Fig. S3 and Table S1 for further data. (**i**) Lineage-tracing strategy during serial regression and regeneration. (**j**–**l**,**n**) YFP$^+$ luminal cells (arrows) that co-express AR (**k**), Nkx3.1 (**l**) and CK18 (**n**) are more frequently observed after three rounds (**j**–**l**) and five rounds (**n**) of serial regeneration. (**m**) Castration-resistant luminal cells can be detected after four rounds of regression. (**o**) The frequency of luminal cells among total YFP$^+$ cells during regeneration in anterior prostate is 0.04% ($n = 5$ out of 11,427, 3 animals) after one round, 0.07% ($n = 13$ out of 18,025, 5 animals) after one round using an alternative protocol, 0.03% ($n = 3$ out of 10,249, 3 animals) after one round in 12-month-old mice, 0.6% ($n = 56$ out of 9,129, 3 animals) after three rounds, and 3.4% ($n = 303$ out of 8,875, 3 animals) after five rounds. The one-round frequencies are not statistically different, whereas $P < 0.0001$ for frequencies of different rounds by χ^2 test. See Supplementary Fig. S4 and Table S1 for further data. AP, anterior prostate; DLP, dorsolateral prostate. Scale bars, 50 μm (**b**–**g**,**j**–**n**). Error bars in **h**,**o** correspond to s.d. and show variability between animals.

YFP$^+$ cells with increasing rounds of regeneration, which is further supported by the detection of castration-resistant YFP$^+$ luminal cells in the regressed prostate after four rounds (Fig. 2m). Taken together, these results are consistent with a model in which a bipotential basal progenitor can give rise to luminal progeny with transit-amplifying characteristics.

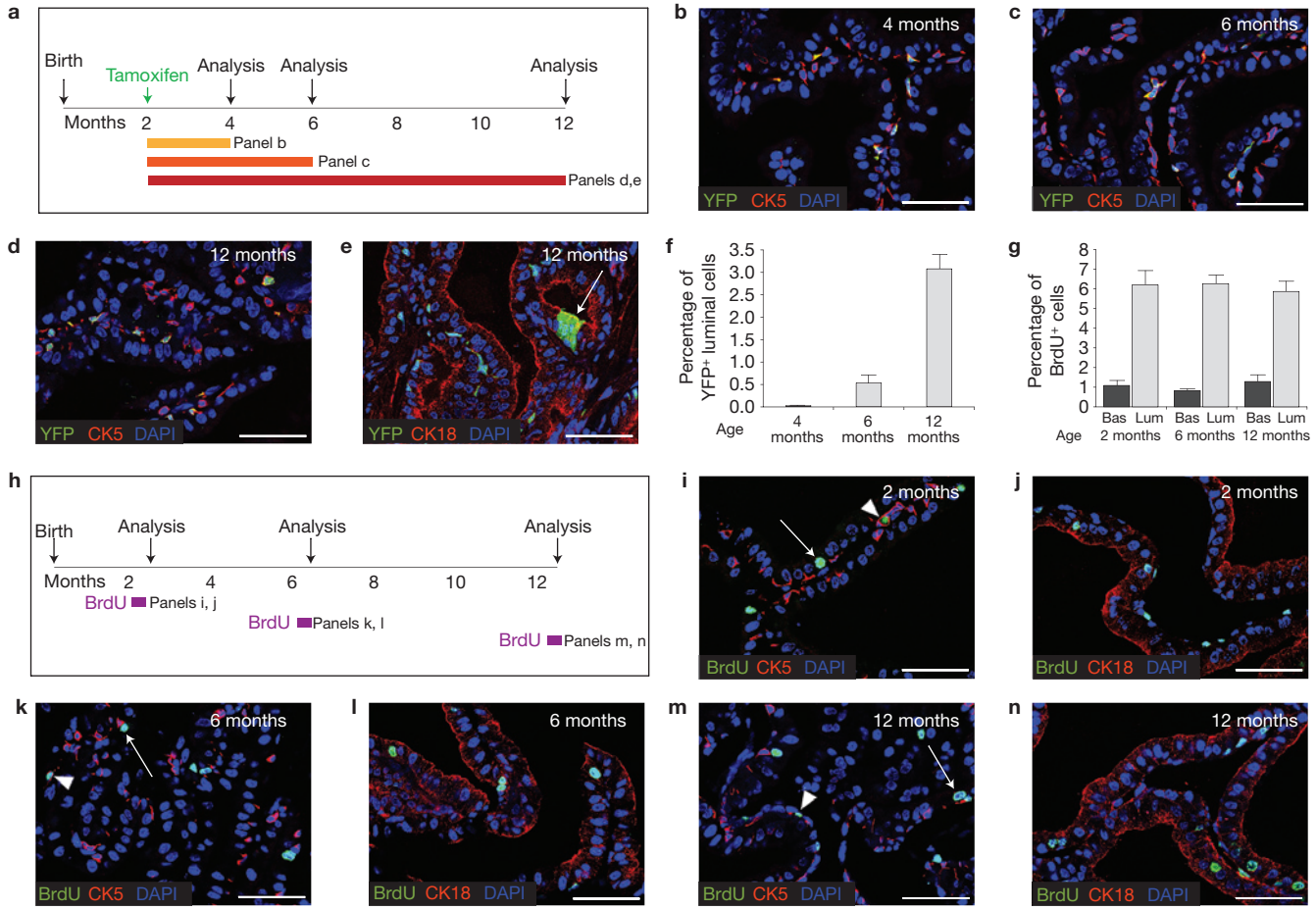

Figure 3 Detection of rare bipotential basal progenitors during prostate homeostasis. (**a**) Time course of lineage-tracing analysis in hormonally intact *CK5–CreER*^{T2} *; R26R–YFP/+* mice. (**b–d**) Co-localization of YFP and CK5 in prostate basal cells in mice at 4 months (**b**), 6 months (**c**) and 12 months (**d**) of age. (**e**) Detection of YFP^{+}CK18^{+} luminal cells (arrow) at 12 months of age. (**f**) The frequency of luminal cells among total YFP^{+}cells during homeostasis is 0.02% ($n = 2/8{,}848$) at 4 months, 0.5% ($n = 57/10{,}572$) at 6 months and 3.0% ($n = 227/7{,}638$) at 12 months; 3 animals were analysed at each time point. $P < 0.0001$ for frequencies at different time points by χ^2 test. (**g**) Graphical summary of BrdU incorporation analyses during homeostasis; 3 animals were analysed for each experiment. BrdU incorporation frequencies at different time points are not statistically different by χ^2 test. See Supplementary Table S1 for further data. (**h**) Strategy for analyses of cell proliferation at three different ages in wild-type C57BL/6 mice. At 2 months, 6 months or 12 months of age, BrdU was administered for 12 days followed by analysis. (**i-n**) Analysis of co-localization of BrdU immunostaining with CK5 (**i**,**k**,**m**) or CK18 (**j**,**l**,**n**); arrowheads indicate BrdU^{+} basal cells, and arrows indicate BrdU^{+} luminal cells. Scale bars, 50 μm (**b–e**,**i–n**). Error bars in **f**,**g** correspond to s.d. and show variability between animals.

Identification of rare bipotential basal cells during prostate homeostasis *in vivo*

We also examined the generation of luminal cells by lineage-marked basal cells during prostate homeostasis, using tamoxifen induction at 2 months of age followed by a long chase period (Fig. 3a). As in our analysis of serial regeneration, we found that the frequency of YFP^{+} luminal cells among all YFP^{+} cells increased with time, as determined at 4 months (0.02%), 6 months (0.5%) or 12 months (3.0%) of age (Fig. 3b–f). However, using BrdU incorporation assays, we found that the proliferation rate in both basal (1.1%) and luminal (6.1%) compartments was relatively stable at 2, 6 and 12 months of age (Fig. 3g–n). These results suggest that low-frequency generation of luminal cells from a bipotential basal progenitor results in small clusters of lineage-marked luminal cells, owing to the higher proliferation rate of luminal cells during homeostasis. Furthermore, these findings indicate that basal cells exhibit bipotentiality during regeneration and tissue homeostasis at approximately similar frequencies.

Basal cells as a cell of origin for prostate cancer

Next, we investigated whether CK5^{+} basal cells can be a cell of origin for prostate cancer, using a conditional allele of the *Pten* tumour suppressor for inducible inactivation in mice, modelling one of the most frequent genetic alterations in human prostate cancer[22,23]. We found that *CK5–CreER*^{T2} *; Pten*$^{flox/flox}$ *; R26R–YFP/+* prostates at 1 month following tamoxifen induction exhibited small foci of epithelial hyperplasia and Grade I prostatic intraepithelial neoplasia (PIN) lesions[24] in otherwise histologically normal glands (Fig. 4a,b and Supplementary Table S2). At this stage, we could detect phosphorylated Akt in basal cells, as expected after loss of *Pten* (Fig. 4f). By 3 months after induction, these mice exhibited Grade II and III PIN lesions, and at 6 months they had Grade III and IV lesions (Fig. 4c,d and Supplementary Fig. S5 and Table S2). Interestingly, we observed increased cell proliferation in basal cells before PIN lesion formation (Fig. 4g and Supplementary Table S2), and subsequently in the PIN/tumour lesions (Fig. 4h). Notably, even

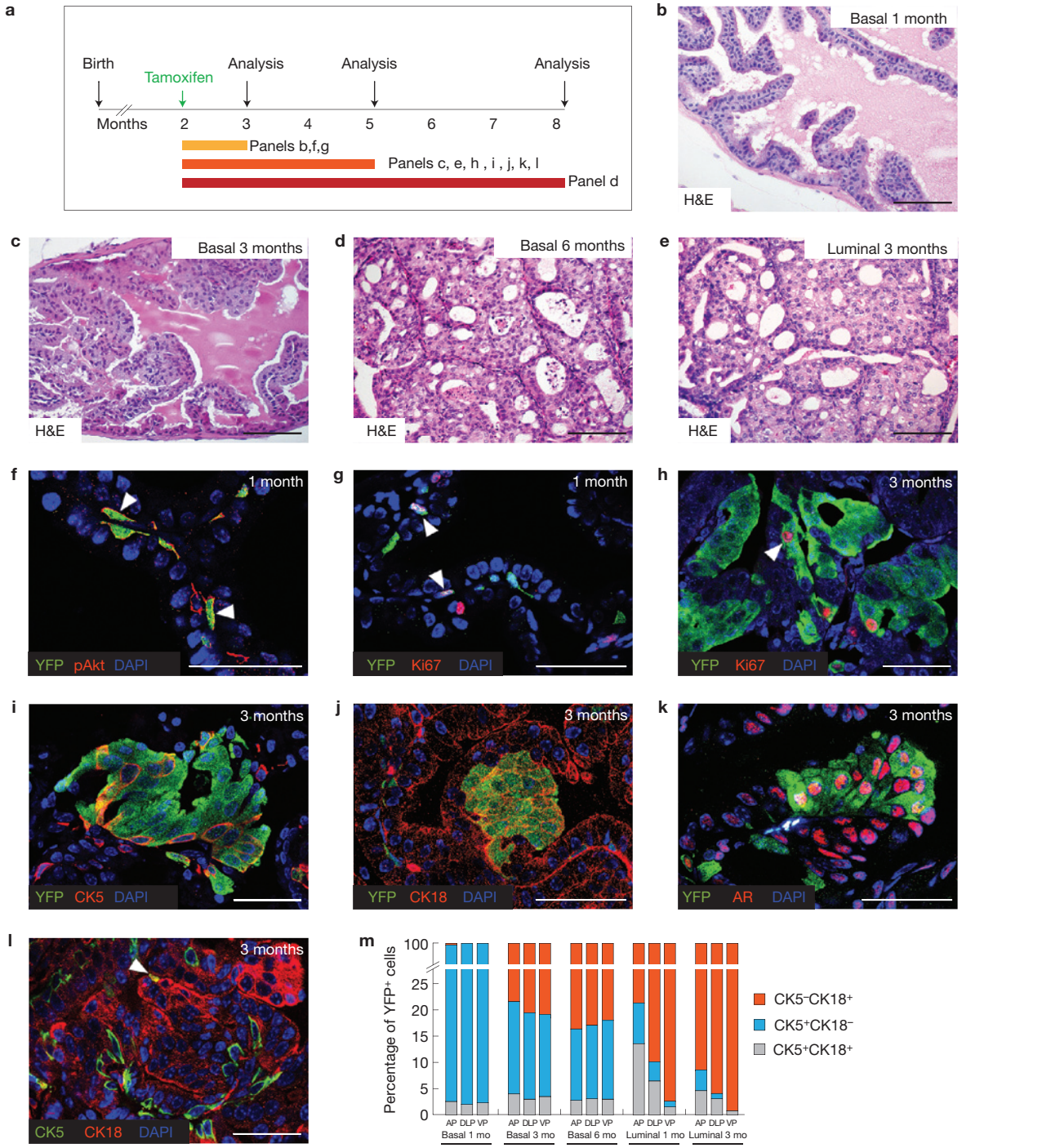

Figure 4 Basal cells are a cell type of origin for prostate tumours. (**a**) Time course for tumour formation in hormonally intact *CK5–CreER*^{T2} ; *Pten*$^{flox/flox}$; *R26R–YFP/+* mice. (**b**–**d**) Haematoxylin–eosin staining of anterior prostates showing slight epithelial hyperplasia at 1 month after induction (**b**), low-grade PIN at 3 months after induction (**c**) and high-grade PIN at 6 months after induction (**d**). (**e**) High-grade PIN in *Nkx3.1*$^{CreERT2/+}$; *Pten*$^{flox/flox}$; *R26R–YFP/+* anterior prostate at 3 months after induction. (**f**-**l**) Marker analysis of PIN lesions in *CK5–CreER*^{T2} ; *Pten*$^{flox/flox}$; *R26R–YFP/+* anterior prostate. (**f**) Phosphorylated Akt (pAkt) can be detected in basal cells at 1 month after induction. (**g**,**h**) Ki67 immunoreactivity can be detected in basal cells at 1 month after induction, before PIN lesion formation (**g**), as well as at 3 months after induction (**h**). (**i**–**k**) Most transformed cells at 3 months after induction do not express CK5 (**i**), but instead express CK18 (**j**) and AR (**k**). (**l**) CK5$^+$ CK18$^+$ intermediate cells (arrowhead) can be detected in PIN lesions at 3 months after induction. (**m**) Quantification of basal (CK5$^+$CK18$^-$), luminal (CK5$^-$CK18$^+$) and intermediate cells (CK5$^+$CK18$^+$) in YFP$^+$prostate cells of *CK5–CreER*^{T2} ; *Pten*$^{flox/flox}$; *R26R–YFP/+* and *Nkx3.1*$^{CreERT2/+}$; *Pten*$^{flox/flox}$; *R26R–YFP/+* mice at the indicated times after induction. Scale bars, 100 μm (**b**–**e**); 50 μm (**f**–**l**).

small PIN lesions contained few CK5$^+$ basal cells (Fig. 4i,m), but instead were mostly comprised of CK18$^+$ and AR$^+$ luminal cells (Fig. 4j,k,m), indicating that oncogenic transformation of basal cells promotes luminal differentiation of their progeny. Furthermore, the low frequency of CK5$^+$ CK18$^+$ intermediate cells in the untransformed prostate (1.6%; Supplementary Fig. S1o–s) and during tumour

formation (3.0%; Fig. 4l,m) suggests that luminal differentiation occurs from transformed basal cells, and does not arise exclusively from intermediate cells.

We next examined whether the time course and histopathology of PIN/tumour lesions in *CK5–CreER^{T2} ; Pten$^{flox/flox}$; R26R–YFP/+* mice differed from those in *Nkx3.1$^{CreERT2/+}$; Pten$^{flox/flox}$; R26R–YFP/+* mice (Fig. 4e and Supplementary Fig. S6a–n), which have a luminal cell of origin. Consistent with a recent report[18], we found that the overall histopathology of the luminal origin PIN/tumour lesions was very similar, but arose with a different time course. At 1 month after induction, the *Nkx3.1$^{CreERT2/+}$; Pten$^{flox/flox}$; R26R–YFP/+* mice exhibited Grade II and III PIN, resembling 3-month basal origin lesions (Fig. 4c and Supplementary Fig. S6b and Table S2). Similarly, 3-month luminal origin tumours exhibited Grade III and IV PIN, resembling 6-month basal origin tumours (Fig. 4d,e and Supplementary Fig. S6c,d and Table S2). These histopathological similarities were further supported by marker analyses (Fig. 4h–k and Supplementary Fig. S6f–k). However, we did note that CK5$^+$ CK18$^+$ cells occurred at low frequencies in all prostate lobes of basal origin lesions, but at higher frequencies in the anterior and dorsolateral lobes of luminal origin lesions (Fig. 4l,m and Supplementary Fig. S6l–n). Overall, basal origin tumours are histologically similar to luminal origin tumours, but arise more slowly, perhaps owing to differences in the starting number of cells undergoing transformation and/or an intrinsic delay due to luminal differentiation from basal cells.

A luminal origin molecular signature predicts patient survival

To determine whether basal and luminal origin tumours might exhibit molecular differences, we performed transcriptome analyses comparing PIN/tumour lesions from basal or luminal origins at time points at which they exhibited similar histopathological phenotypes. We used RNA-seq to profile prostate tissue from basal origin tumours of *CK5–CreER^{T2} ; Pten$^{flox/flox}$; R26R–YFP/+* mice that were uninduced (control), or at 3 or 6 months after tamoxifen induction, or from luminal origin tumours of *Nkx3.1$^{CreERT2/+}$; Pten$^{flox/flox}$; R26R–YFP/+* mice that were uninduced (control), or at 1 or 3 months after induction (6 mice per category). Principal components analysis demonstrated the reproducibility of the independent biological replicates (Fig. 5a,b). We then performed gene set enrichment analysis (GSEA) to compare initiation signatures for basal origin lesions (3 months versus control) and for luminal origin lesions (1 month versus control; Supplementary Tables S3 and S4); a similar comparison was performed for progression signatures for basal origin (6 months versus 3 months) and luminal origin lesions (3 months versus 1 month) (Supplementary Tables S5 and S6). These comparisons demonstrated the strong reciprocal enrichment of the initiation as well as progression signatures (Fig. 5c,d), indicating that basal and luminal origin tumours are globally similar at the molecular level.

Nonetheless, we could identify molecular differences between the luminal and basal origin signatures by applying bioinformatic subtraction of the similar components of their signatures. Thus, we generated a mouse expression signature containing genes upregulated in luminal origin lesions relative to basal origin lesions (luminal 3 months–basal 6 months), or conversely upregulated in basal origin lesions (Supplementary Table S7). Next, we used a mouse-to-human cross-species approach to compare this mouse luminal versus basal signature to a human lethality signature generated using survival data in a Swedish watchful-waiting patient cohort[25]. We defined this lethality signature as a list of genes ranked by their differential expression between high-risk (death within 12 months, 6 samples) and low-risk (survival for more than 192 months, 12 samples) cases in this cohort (Supplementary Table S8). Using GSEA, we found that genes overexpressed in luminal origin tumours were significantly enriched in genes upregulated in the lethality (high-risk) signature (Fig. 5e). In contrast, genes overexpressed in the basal origin tumours exhibited no statistical enrichment in the human lethality signature (Fig. 5f), suggesting that luminal origin prostate cancer is more aggressive than basal origin prostate cancer. Furthermore, we used the luminal versus basal origin signature to infer Gene Ontology-Biological Process (GO-BP) gene sets that were significantly upregulated or downregulated. We identified 57 such GO-BP categories ($P < 0.05$; Fig. 5g and Supplementary Table S9), including categories such as cell cycle mitotic, E2F-mediated regulation of DNA replication, and extension of telomeres, consistent with a more aggressive phenotype of luminal origin tumours.

Next, we investigated the clinical relevance of the 68 genes with significant contribution to the enrichment of the luminal origin upregulated signature in the human lethality signature (the leading edge in Fig. 5e and Supplementary Table S10). Interestingly, many of these 68 genes are upregulated in human prostate cancer, as shown by analysis of five human patient data sets[22,26–29] (Supplementary Fig. S6o and Table S10). We tested this luminal origin leading-edge signature (LOLES) with two different data sets from Memorial Sloan-Kettering Cancer Center[22,30], in which patients are characterized by their biochemical recurrence (BCR)-free survival time, representing the duration between prostatectomy and subsequent detection of rising serum prostate-specific antigen levels. Using k-means clustering, we could stratify the 79 primary tumour samples from the Glinsky data set[30], and the 131 primary tumour samples from the Taylor data set[22], into two groups with significant differences in BCR-free survival (log-rank $P = 0.0255$, Fig. 5h; log-rank $P = 0.0245$, Fig. 5i). Furthermore, the LOLES successfully stratified 263 patients of the Swedish cohort (excluding the samples used to define the lethality signature) into two distinct groups with a statistically significant difference in mean survival time of 3.4 years at 50% survival (log-rank $P = 0.0065$, Fig. 5j).

Finally, we examined whether the LOLES may have independent prognostic value compared with histological Gleason scoring, which remains the best prognostic marker for overall survival[31]. Indeed, C-statistics analysis using the Swedish watchful-waiting cohort revealed that the LOLES improves the prognostic value of the Gleason score from 0.76 ($C = 0.76$; 95% CI 0.72–0.80, $P = 1.10 \times 10^{-39}$) to 0.82 ($C = 0.82$, 95% CI 0.77–0.84, $P = 2.25 \times 10^{-80}$; Fig. 5k). Taken together, our bioinformatic analysis shows that the LOLES is highly correlated with poor patient prognosis.

DISCUSSION

Using a lineage-marking approach to examine the same population in *ex vivo* and *in vivo* assays, we have shown that prostate basal cells exhibit distinct properties in different assays for stem cell function, and reconcile several aspects of previous work on prostate stem cells that have seemed to be discordant. Notably, studies using *ex vivo* cell culture and tissue reconstitution assays have

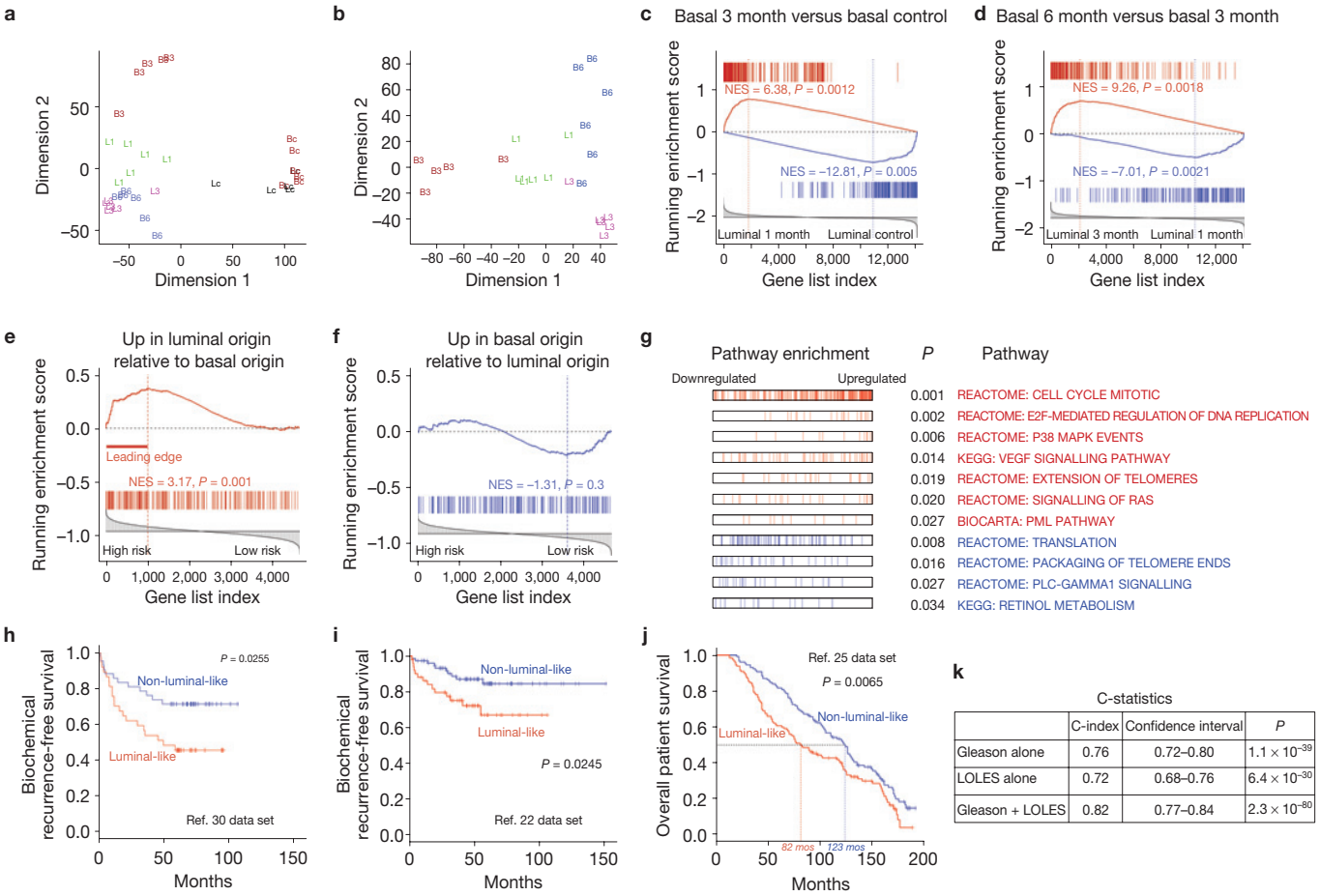

Figure 5 A luminal origin gene signature that is prognostic for human prostate cancer outcome. (**a**) Scatter plot of the two main components from a principal components analysis based on 14,063 genes, capturing 55% (dimension 1) and 21% (dimension 2) of the data variability. (**b**) Scatter plot without the control samples. Bc, basal origin control; B3, basal origin 3 months post-induction; B6, basal origin 6 months post-induction; Lc, luminal origin control; L1, luminal origin 1 month post-induction; L3, luminal origin 3 months post-induction. (**c**) GSEA comparison of basal origin initiation signature (B3 relative to Bc) to luminal initiation signature (L1 relative to Lc) shows strong enrichment in both directions. (**d**) GSEA comparison of basal origin progression signature (B6 relative to B3) to luminal progression signature (L3 relative to L1) also shows strong enrichment in both directions. (**e**,**f**) GSEA shows that genes upregulated in the L3 versus B6 gene signature are strongly enriched in a human signature corresponding to lethality due to prostate cancer (**e**), but not the converse (**f**). (**g**) GSEA shows biological pathways significantly enriched in the L3 versus B6 gene signature; *P* is estimated using 1,000 sample permutations. Pathways in red are upregulated in luminal origin tumours; pathways in blue are upregulated in basal origin tumours. (**h**,**i**) Kaplan–Meier analysis shows that the LOLES (corresponding to the 68 genes to the left of the dashed line in **e**) stratifies patients from 2 independent cohorts into groups with different rates of biochemical recurrence (red curve, luminal-like group, 37 patients in **h**, 52 patients in **i**; blue curve, non-luminal-like group, 42 patients in **h**, 79 patients in **i**). (**j**) Kaplan–Meier analysis shows that the LOLES stratifies patients from the Swedish watchful waiting cohort into a luminal-like group (red curve, 132 patients) and a non-luminal-like group (blue curve, 131 patients) with different survival outcomes (82 months versus 123 months at 50% survival). (**k**) C-statistics analysis shows that the LOLES improves the prognostic value of the Gleason score in the Swedish cohort from 0.76 to 0.82, with the 95% confidence intervals and *P* values shown.

identified prostate basal cells as stem cells[8,12,21], but our analyses indicate that basal cells can exhibit substantial plasticity when removed from their endogenous tissue microenvironment. More generally, our work supports the notion that genetic lineage tracing *in vivo*, not transplantation-based assays, represents a gold standard for identification of physiologically relevant stem cells[32], and that unexpected plasticity should be considered when interpreting the outcomes of other stem cell assays.

In principle, the plasticity of prostate basal cells might be regulated by extrinsic or intrinsic factors. For example, stromal and/or luminal cells could inhibit basal cell plasticity, and *Pten* loss might confer independence from this inhibition, allowing formation of transformed luminal cells. Moreover, the embryonic urogenital mesenchyme employed in tissue reconstitution assays has potent reprogramming activity[33,34], and could perhaps reprogram adult prostate basal cells to an embryonic multipotent progenitor state. Another possibility is that basal cell plasticity may be intrinsically regulated by cellular proliferation, because the percentage of luminal progeny is approximately 1–2% of dividing basal cells during regeneration and adult homeostasis. Thus, in response to strong proliferative signals provided by embryonic urogenital mesenchyme in tissue reconstitution assays, or due to *Pten* inactivation, basal cells might generate increased numbers of luminal cells.

Our findings suggest an overall similarity in the functional properties of basal cells in the prostate to those in other ductal epithelial tissues. Recent lineage-tracing analyses of the mammary and sweat gland

epithelium have shown that luminal and myoepithelial cells are maintained by distinct unipotent progenitors, and myoepithelial cells can generate luminal cells in fad pad transplantation assays[35,36]. Our work, together with other recent studies[18,37], indicates that a similar lineage relationship exists in the prostate epithelium, and that the plasticity of myoepithelial/basal cells may be a conserved attribute of ductal epithelial tissues. Moreover, the plasticity of basal cells in the adult prostate epithelium may reflect their role as stem/progenitor cells that generate all epithelial cell types during organogenesis[38], similar to basal progenitors in mammary and sweat gland development[35,36].

We have also detected rare prostate basal cells that exhibit bipotentiality during adult tissue regeneration and homeostasis, at a frequency similar to that of bipotential luminal CARNs in the regressed prostate[14]. This finding differs from the conclusions of a recent report that the prostate basal lineage is completely unipotent[18]. The basis for this discrepancy is unclear, but could be due to the different Cre drivers used, or the larger number of cells and rounds of regeneration analysed in this study. Notably, the existence of bipotential adult basal stem/progenitor cells is also supported by retrospective lineage analyses of human prostate tissue using mitochondrial DNA mutations[39,40]. Taken together, our findings suggest that basal and luminal lineages are largely independent in the adult prostate, but rare basal and luminal stem/progenitor cells can potentially compensate for imbalances in cell number during regeneration and tissue homeostasis. One possible model is that rare basal (and luminal) stem/progenitor cells may reside at the top of an epithelial lineage hierarchy, and a larger subpopulation of basal cells, perhaps corresponding to transit-amplifying cells, can exhibit plasticity in sphere and tissue reconstitution assays (Fig. 6a). Alternatively, basal cells may exhibit stochastic stem/progenitor properties, with a low probability in the adult prostate epithelium and a higher probability when explanted or transformed (Fig. 6b), resembling the maintenance of interfollicular and esophageal epithelium by stochastic progenitors during homeostasis and wound repair[41,42].

In humans, prostate adenocarcinoma exhibits a strong luminal phenotype with relatively uniform histopathological characteristics, and lacks distinct histological subtypes. At the molecular level, there has been some success in classifying tumours on the basis of gene expression profiling[26,43], and distinct molecular subtypes may be identifiable by specific mutations and/or chromosomal rearrangements[22,23,44,45]. Nonetheless, prostate cancer has previously seemed to differ from other human cancers in which distinct tumour subtypes are readily defined[16,46–48]. However, although we find that basal and luminal origin tumours in mice are histologically similar, as also reported previously[18], our mouse-to-human cross-species bioinformatic analysis has identified a molecular signature in luminal origin tumours that correlates with patient outcome. Thus, our analysis suggests that prostate tumours arising from different cell types of origin may have distinct prognostic outcomes and/or treatment responses.

A major clinical challenge in prostate cancer research has been to distinguish the minority of patients who will develop aggressive disease from those who have indolent cancer and require minimal treatment. Indeed, considerable variation in outcome can exist between tumours with identical Gleason scores[31], indicating that intertumour heterogeneity exists among prostate cancers that are histologically indistinguishable. So far, it has been difficult to identify useful prognostic biomarkers to improve on Gleason scoring, despite considerable effort to identify molecular signatures that can successfully stratify aggressive from non-aggressive disease[25]. Our cross-species analyses suggest that the cell of origin may represent an important component in determining prostate cancer aggressiveness, and may therefore lead to the identification of successful biomarkers. □

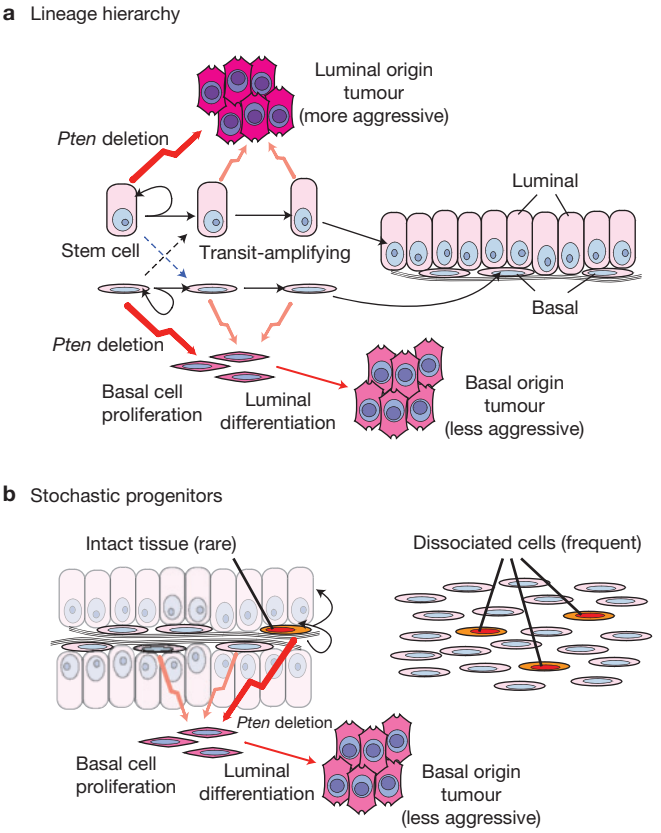

Figure 6 Two models for prostate epithelial lineage relationships and cell of origin for cancer. (**a**) In a conventional lineage hierarchy model, luminal and basal lineages are independently maintained by largely unipotent stem/progenitor cells in the normal adult prostate epithelium. However, luminal and basal progenitors can generate the other cell type during prostate regeneration and tissue homeostasis (dashed lines); in the case of luminal stem/progenitor cells, it remains unclear whether such bipotentiality (blue dashed line) is exhibited only by CARNs in the regressed state. In the case of the basal lineage, bipotential stem cells are relatively rare (approximately 0.05%), whereas basal cells that can exhibit stem cell properties in sphere formation and tissue reconstitution assays are more common (approximately 4%), and perhaps might correspond to transit-amplifying cells. Oncogenic transformation of either luminal or basal cells by inactivation of *Pten* results in tumours with histologically similar luminal phenotypes, but tumours arising from basal cells first undergo basal cell proliferation and subsequently luminal differentiation. Tumours may arise from stem cells (dark red jagged arrows) or may also be derived from more differentiated cell types (light red jagged arrows). (**b**) In a stochastic progenitor model, basal cells within an intact prostate epithelium randomly exhibit stem/progenitor properties at very low frequencies (orange), giving rise to luminal cells and being capable of self-renewal. (Luminal cells could conceivably follow a similar stochastic progenitor model, but this is not shown.) After tissue dissociation, however, the probability of such random basal stem/progenitor cells may be greatly increased. Oncogenic transformation of normal basal cells and/or stochastic basal progenitors leads to luminal differentiation and tumour formation; however, it is unlikely that stochastic progenitors represent the sole cell of origin following *Pten* deletion, given the rarity of these progenitors versus the frequency of observed PIN lesions.

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Supplementary Information is available in the online version of the paper

ACKNOWLEDGEMENTS

We thank J. Rock and B. Hogan (Duke University Medical School, USA) for their generous gift of *CK5–CreER*^{T2} transgenic mice, O. Couronne and the Columbia University Genome Sequencing Facility for RNA-seq analysis, and R. DePinho (MD Anderson Cancer Center) for assistance in obtaining the Glinsky data set. This work was supported by a postdoctoral fellowship from the DOD Prostate Cancer Research Program (Z.A.W.), a postdoctoral Computing Innovation Fellowship (CIFellow) from NSF, CRA and CCC (A.M.), and by grants from the National Institutes of Health (C.A-S., A.C. and M.M.S.).

AUTHOR CONTRIBUTIONS

Z.A.W. carried out mouse experiments, Z.A.W. and S.K.B. performed transplantation experiments, A.M. performed bioinformatic analyses, and R.D.C. analysed tumour histology. Z.A.W. and M.M.S. designed the overall study, C.A-S., R.D.C., A.C. and M.M.S. supervised the data analysis, and Z.A.W., A.M. and M.M.S. wrote the manuscript, in consultation with C.A-S. and A.C. All authors provided discussion and comments on the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at www.nature.com/doifinder/10.1038/ncb2697
Reprints and permissions information is available online at www.nature.com/reprints

1. Visvader, J. E. Cells of origin in cancer. *Nature* **469,** 314–322 (2011).
2. Shen, M. M. & Abate-Shen, C. Molecular genetics of prostate cancer: new prospects for old challenges. *Genes Dev.* **24,** 1967–2000 (2010).
3. Goldstein, A. S., Stoyanova, T. & Witte, O. N. Primitive origins of prostate cancer: *In vivo* evidence for prostate-regenerating cells and prostate cancer-initiating cells. *Mol. Oncol.* **4,** 385–396 (2010).
4. Moscatelli, D. & Wilson, E. L. PINing down the origin of prostate cancer. *Sci. Transl. Med.* **2,** 43ps38 (2010).
5. Wang, Z. A. & Shen, M. M. Revisiting the concept of cancer stem cells in prostate cancer. *Oncogene* **30,** 1261–1271 (2011).
6. De Marzo, A. M., Nelson, W. G., Bieberich, C. J. & Yegnasubramanian, S. Prostate cancer: new answers prompt new questions regarding cell of origin. *Nat. Rev. Urol.* **7,** 650–652 (2010).
7. Taylor, R. A., Toivanen, R. & Risbridger, G. P. Stem cells in prostate cancer: treating the root of the problem. *Endocr. Relat. Cancer* **17,** R273–R285 (2010).
8. Lawson, D. A., Xin, L., Lukacs, R. U., Cheng, D. & Witte, O. N. Isolation and functional characterization of murine prostate stem cells. *Proc. Natl Acad. Sci. USA* **104,** 181–186 (2007).
9. Richardson, G. D. *et al.* CD133, a novel marker for human prostatic epithelial stem cells. *J. Cell Sci.* **117,** 3539–3545 (2004).
10. Burger, P. E. *et al.* Sca-1 expression identifies stem cells in the proximal region of prostatic ducts with high capacity to reconstitute prostatic tissue. *Proc. Natl Acad. Sci. USA* **102,** 7180–7185 (2005).
11. Xin, L., Lawson, D. A. & Witte, O. N. The Sca-1 cell surface marker enriches for a prostate-regenerating cell subpopulation that can initiate prostate tumorigenesis. *Proc. Natl Acad. Sci. USA* **102,** 6942–6947 (2005).
12. Goldstein, A. S. *et al.* Trop2 identifies a subpopulation of murine and human prostate basal cells with stem cell characteristics. *Proc. Natl Acad. Sci. USA* **105,** 20882–20887 (2008).
13. Goldstein, A. S., Huang, J., Guo, C., Garraway, I. P. & Witte, O. N. Identification of a cell of origin for human prostate cancer. *Science* **329,** 568–571 (2010).
14. Wang, X. *et al.* A luminal epithelial stem cell that is a cell of origin for prostate cancer. *Nature* **461,** 495–500 (2009).
15. Shibata, M. & Shen, M. M. The roots of cancer: stem cells and the basis for tumor heterogeneity. *Bioessays* http://dx.doi.org/10.1002/bies.201200101 (2012).
16. Visvader, J. E. Keeping abreast of the mammary epithelial hierarchy and breast tumorigenesis. *Genes Dev.* **23,** 2563–2577 (2009).
17. Lawson, D. A. *et al.* Basal epithelial stem cells are efficient targets for prostate cancer initiation. *Proc. Natl Acad. Sci. USA* **107,** 2610–2615 (2010).
18. Choi, N., Zhang, B., Zhang, L., Ittmann, M. & Xin, L. 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).
19. Rock, J. R. *et al.* Basal cells as stem cells of the mouse trachea and human airway epithelium. *Proc. Natl Acad. Sci. USA* **106,** 12771–12775 (2009).
20. Srinivas, S. *et al.* Cre reporter strains produced by targeted insertion of EYFP and ECFP into the ROSA26 locus. *BMC Dev. Biol.* **1,** 4 (2001).
21. Xin, L., Lukacs, R. U., Lawson, D. A., Cheng, D. & Witte, O. N. Self-renewal and multilineage differentiation *in vitro* from murine prostate stem cells. *Stem Cells* **25,** 2760–2769 (2007).
22. Taylor, B. S. *et al.* Integrative genomic profiling of human prostate cancer. *Cancer Cell* **18,** 11–22 (2010).
23. Berger, M. F. *et al.* The genomic complexity of primary human prostate cancer. *Nature* **470,** 214–220 (2011).
24. Park, J. H. *et al.* Prostatic intraepithelial neoplasia in genetically engineered mice. *Am. J. Pathol.* **161,** 727–735 (2002).
25. Sboner, A. *et al.* Molecular sampling of prostate cancer: a dilemma for predicting disease progression. *BMC Med. Genom.* **3,** 8 (2010).
26. Lapointe, J. *et al.* Gene expression profiling identifies clinically relevant subtypes of prostate cancer. *Proc. Natl Acad. Sci. USA* **101,** 811–816 (2004).
27. Tomlins, S. A. *et al.* Integrative molecular concept modeling of prostate cancer progression. *Nat. Genet.* **39,** 41–51 (2007).
28. Singh, D. *et al.* Gene expression correlates of clinical prostate cancer behavior. *Cancer Cell* **1,** 203–209 (2002).
29. Yu, Y. P. *et al.* Gene expression alterations in prostate cancer predicting tumor aggression and preceding development of malignancy. *J. Clin. Oncol.* **22,** 2790–2799 (2004).
30. Glinsky, G. V., Glinskii, A. B., Stephenson, A. J., Hoffman, R. M. & Gerald, W. L. Gene expression profiling predicts clinical outcome of prostate cancer. *J. Clin. Invest.* **113,** 913–923 (2004).
31. Martin, N. E., Mucci, L. A., Loda, M. & Depinho, R. A. Prognostic determinants in prostate cancer. *Cancer J.* **17,** 429–437 (2011).
32. Snippert, H. J. & Clevers, H. Tracking adult stem cells. *EMBO Rep.* **12,** 113–122 (2011).
33. Taylor, R. A. *et al.* Lineage enforcement by inductive mesenchyme on adult epithelial stem cells across developmental germ layers. *Stem Cells* **27,** 3032–3042 (2009).
34. Cunha, G. R. *et al.* The endocrinology and developmental biology of the prostate. *Endocr. Rev.* **8,** 338–362 (1987).
35. Van Keymeulen, A. *et al.* Distinct stem cells contribute to mammary gland development and maintenance. *Nature* **479,** 189–193 (2011).
36. Lu, C. P. *et al.* Identification of stem cell populations in sweat glands and ducts reveals roles in homeostasis and wound repair. *Cell* **150,** 136–150 (2012).
37. Liu, J. *et al.* Regenerated luminal epithelial cells are derived from preexisting luminal epithelial cells in adult mouse prostate. *Mol. Endocrinol.* **25,** 1849–1857 (2011).
38. Ousset, M. *et al.* Multipotent and unipotent progenitors contribute to prostate postnatal development. *Nat. Cell Biol.* **14,** 1131–1138 (2012).
39. Blackwood, J. K. *et al. In situ* lineage tracking of human prostatic epithelial stem cell fate reveals a common clonal origin for basal and luminal cells. *J. Pathol.* **225,** 181–188 (2011).
40. Gaisa, N. T. *et al.* Clonal architecture of human prostatic epithelium in benign and malignant conditions. *J. Pathol.* **225,** 172–180 (2011).
41. Doupe, D. P. *et al.* A single progenitor population switches behavior to maintain and repair esophageal epithelium. *Science* **337,** 1091–1093 (2012).
42. Doupe, D. P., Klein, A. M., Simons, B. D. & Jones, P. H. The ordered architecture of murine ear epidermis is maintained by progenitor cells with random fate. *Dev. Cell* **18,** 317–323 (2010).
43. Markert, E. K., Mizuno, H., Vazquez, A. & Levine, A. J. Molecular classification of prostate cancer using curated expression signatures. *Proc. Natl Acad. Sci. USA* **108,** 21276–21281 (2011).
44. Lapointe, J. *et al.* Genomic profiling reveals alternative genetic pathways of prostate tumorigenesis. *Cancer Res.* **67,** 8504–8510 (2007).
45. Rubin, M. A., Maher, C. A. & Chinnaiyan, A. M. Common gene rearrangements in prostate cancer. *J. Clin. Oncol.* **29,** 3659–3668 (2011).
46. Alizadeh, A. A. *et al.* Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* **403,** 503–511 (2000).
47. Gibson, P. *et al.* Subtypes of medulloblastoma have distinct developmental origins. *Nature* **468,** 1095–1099 (2010).
48. Collisson, E. A. *et al.* Subtypes of pancreatic ductal adenocarcinoma and their differing responses to therapy. *Nat. Med.* **17,** 500–503 (2011).

DOI: 10.1038/ncb2697

METHODS

Mouse strains and genotyping. The *Nkx3.1*^{CreERT2} targeted allele[14] and *CK5–CreER*^{T2} transgenic line[19] have been described previously. Genotyping was performed by PCR using tail genomic DNA, with the following primer sequences: *Nkx3.1* wild-type allele, 5′-CTCCGCTACCCTAAGCATCC-3′ and 5′-GACACTGTCATATTACTTGGACC-3′; *CreER*^{T2} allele, 5′-CAGATGGCGCGGC-AACACC-3′ and 5′-GCGCGGTCTGGCAGTAAAAAC-3′; *Pten*^{flox} allele, 5′-ACTCAAGGCAGGGATGAGC-3′ and 5′-GTCATCTTCACTTAGCCATTGG-3′; *R26R–YFP* allele, 5′-GCGAAGAGTTTGTCCTCAACC-3′ (mutated forward), 5′-GGAGCGGGAGAAATGGATATG-3′ (wild-type forward) and 5′-AAAGTCGCTCTGAGTTGTTAT-3′ (wild-type and mutated reverse).

Mouse procedures. Castration of adult male mice was performed using standard techniques, with the fully regressed state attained at 4 weeks after castration. For prostate regeneration, testosterone (Sigma) was dissolved at 25 mg ml^{-1} in 100% ethanol and diluted in PEG-400 to a final concentration of 7.5 mg ml^{-1}. Testosterone was administered for 4 weeks at a rate of 1.875 µg h^{-1} delivered by subcutaneous implantation of mini-osmotic pumps (Alzet). This regimen yields physiological levels of serum testosterone[49].

For tamoxifen induction, mice were administered 9 mg tamoxifen (Sigma) suspended in corn oil per 40 g body weight, or vehicle alone for negative controls, by oral gavage once daily for 4 consecutive days, followed by a chase period of 14 days. BrdU (100 mg kg^{-1}; Sigma) was administered by intraperitoneal injection twice daily for 12 consecutive days during regeneration or homeostasis to label proliferating cells.

Tissue collection and flow cytometry. For histological and immunofluorescence analysis, individual prostate lobes or renal grafts were dissected and fixed in 4% paraformaldehyde for subsequent cryoembedding in OCT compound (Sakura), or fixed in 10% formalin followed by paraffin embedding. For RNA isolation and RNA-seq analysis, prostate tissues were quickly dissected, flash-frozen in liquid nitrogen and stored at −80 °C.

For flow cytometry, prostate tissues were dissected and minced to small clumps, followed by enzymatic dissociation with 0.2% collagenase I (Invitrogen) in DMEM media with 5% FBS for 3 h at 37 °C. Tissues were digested with 0.25% trypsin–EDTA (StemCell Technologies) for 1 h at 4 °C, passed through 21- to 26-gauge syringes and filtered through a 40-µm cell strainer to obtain single-cell suspensions. Dissociated prostate cells were suspended in Hanks' balanced salt solution modified/2% FBS. Cell sorting was performed on a BD FACS Aria II instrument in the Flow Cytometry Shared Resource of the Herbert Irving Comprehensive Cancer Center. Antibodies used for sorting of Lin^{-}Sca-1^{+}CD49f^{hi} cells are listed in Supplementary Table S11.

Sphere formation assay. For sphere formation assays, dissociated cells were incubated in PrEGM medium (Lonza). For each sample, 40 µl of cell suspension was mixed with 60 µl cold Matrigel, and pipetted around the rim of a well of a 12-well plate. The plates were placed in a 37 °C CO_2 incubator for 30 min to allow the Matrigel to solidify. Warm PrEGM (800 µl) was then added to each well. The spheres were cultured and monitored for 7–10 days with 50% medium change every 3 days. For sphere differentiation experiments, the spheres were subsequently cultured in PrEGM medium with 10^{-8} M dihydrotestosterone (DHT), and monitored for 6 days with 50% medium change every 2 days.

Tissue reconstitution assay. For limiting dilution analysis, 5,000, 100 or 20 dissociated YFP^{+} cells obtained from tamoxifen-induced *CK5–Cre*^{ERT2} ; *R26R–YFP*/+ mice were mixed with 2.5 × 10^{5} dissociated urogenital sinus mesenchyme (UGM) cells from embryonic day 18.0 rat embryos. UGM cells were obtained from dissected urogenital sinus treated for 30 min in 1% trypsin, followed by mechanical dissociation and treatment with 0.1% collagenase B (Roche) for 30 min at 37 °C, and washing in PBS. Pelleted cell mixtures were resuspended in 10 µl of 1:8 collagen/setting buffer (10× Earle's Balanced Salt Solution (Life Technologies), 0.2 M $NaHCO_3$ and 50 mM NaOH), and gelatinized at 37 °C for 20 min. Tissue recombinants were cultured in DMEM/10% FBS supplemented with 10^{-7} M DHT overnight, followed by transplantation under the kidney capsules of immunodeficient NOD.Cg-*Prkdc*^{scid} *Il2rg*^{tm1Wjl}/SzJ (NSG) mice (Jackson Laboratory). Grafts were collected after 8–12 weeks of growth for analysis, and extreme limiting dilution analysis was performed as described previously[50].

Histology and immunostaining. Haematoxylin–eosin staining was performed using standard protocols on 6 µm paraffin sections. Histological assessments were performed using a published classification of mouse PIN lesions[24].

For immunohistochemical staining, 6 µm paraffin sections were deparaffinized in xylene, followed by boiling in antigen unmasking solution (Vector Labs). Slides were blocked in 10% normal goat serum (NGS; Vector Labs), and incubated with primary antibodies diluted in 10% NGS overnight at 4 °C. Secondary antibodies were obtained from Vectastain ABC kits (Vector Labs) and diluted 1:250. Signal was enhanced using the Vectastain ABC system and visualized with the NovaRed Substrate Kit (Vector Labs). Slides were counterstained with Harris modified haematoxylin (1:4 diluted in H_2O; Fisher Scientific) and mounted with Clearmount (American MasterTech). Haematoxylin–eosin staining and immunohistochemical staining were imaged using a Nikon Eclipse E800 microscope equipped with a Nikon DXM1200 digital camera.

Immunofluorescence staining was performed using 6 µm cryosections, which were incubated in 3% H_2O_2 and Antigen Unmasking Solution (Vector Labs), or culture plates for whole-mount staining of spheres fixed in 4% paraformaldehyde for 15 min. Samples were incubated with 10% NGS and primary antibodies diluted in 10% NGS overnight at 4 °C. Samples were then incubated with secondary antibodies (diluted 1:500 in PBST) labelled with Alexa Fluor 488, 555 or 647 (Invitrogen/Molecular Probes). Detection of Nkx3.1 was enhanced using tyramide amplification (Invitrogen/Molecular Probes) by incubation of slides with HRP-conjugated secondary antibody (1:100 dilution; Invitrogen/Molecular Probes), followed by incubation with tyramide 555 for 6 min. Slides were mounted with VECTASHIELD mounting medium with DAPI (Vector Labs). Immunofluorescence staining was imaged using a Leica TCS SP5 spectral confocal microscope. All primary antibodies and dilutions used are listed in Supplementary Table S11.

Cell numbers were counted manually using confocal ×40 and ×63 photomicrographs. Statistical analyses were performed using a two-sample *t*-test, χ^2 test or Fisher's exact test as appropriate. At least 3 animals for each experiment or genotype were analysed.

RNA sequencing. Total RNA from prostate tissues was isolated using the Nucleospin RNA II kit (Clontech). The quantity and quality of each sample were measured using an Agilent 2100 Bioanalyzer. RNA-seq analysis was performed by the Columbia Genomic Sequencing Core Facility. Fragments per kilobase of transcript per million mapped reads (FPKM) values for 22,310 genes were reported. Genes with missing values (count of 0) in more than 10% of samples were eliminated from the analysis. Missing values for the remaining genes were estimated by using impute.knn (impute package in R v2.11.1). The resulting data set of 14,063 genes was normalized by using the robust spline normalization (RSN) function of R-system v2.11.1 and was log transformed. Principal components analysis was performed on scaled data, where the data value was adjusted by subtracting its mean across all samples and dividing by its standard deviation, $z = (x - \text{mean})/\text{s.d.}$ Expression data are deposited in the Gene Expression Omnibus database under GSE39509.

Statistical data analysis. Differential expression was estimated using the Welch *t*-test (*t*.test function in R v2.11.1). To compare two distinct signatures we used GSEA (ref. 51), where *P* value was estimated with 1,000 sample permutations. Fold-change analysis was performed on data regenerated by reverse log transformation. Kaplan–Meier curves were generated using the surv, survdiff and survfit functions from package survival in R. C-statistics analysis was conducted using the concordance index function from the R survcomp package. The multivariate Cox proportional hazards model analysis with the Gleason score and leading edge genes was computed using a multiplicative and additive model for defining an integrated risk score model.

Comparing initiation and progression genes for luminal and basal tumours. To compare genes that are responsible for initiation in tumours of both origins, we compared the L1 versus Lc signature with the B3 versus Bc signature. The query signature was defined as a list of genes ranked by their differential expression in the B3 group ($n = 5$ samples) compared with the Bc group ($n = 6$ samples) and was divided into two tails: a positive tail containing the top 200 overexpressed genes in B3 compared with Bc, and a negative tail with the top 200 underexpressed genes in B3 compared with Bc. The target signature was defined between L1 ($n = 6$ samples) and Lc ($n = 5$ samples) groups (Supplementary Tables S3 and S4). Statistical significance of the enrichment between the query signatures and the target signatures was computed separately for the positive and negative tails using GSEA. Analogous analyses were performed comparing L3 versus L1 with B6 versus B3 signatures (Supplementary Tables S5 and S6).

Comparing a luminal versus basal origin mouse signature with a human lethality signature. The luminal versus basal mouse signature (Fig. 5e,f) was defined between L3 ($n = 6$ samples) and B6 groups ($n = 6$ samples). For comparison with human prostate cancer expression profiling data, mouse genes were mapped to their human orthologues using mouse–human orthologous relations from the Mouse Genome Informatics database (MGI, http://www.informatics.jax.org/). This resulted in reduction of 14,063 mouse genes and 6,144 human genes to the set of 4,629 genes common between the human and mouse platforms.

The human lethality risk signature was derived from the Swedish watchful-waiting data set[25], consisting of 281 prostate cancer samples from the cohort recruited in Sweden between 1977 and 1999. We defined the high-risk (aggressive) group as patients that survived less than 12 months ($n = 6$) and the low-risk (indolent) group as patients that survived for more than 192 months ($n = 12$).

Genes differentially expressed between mouse L3 and B6 tumours, P value 0.01, were compared to the human lethality signature using GSEA. (Genes that were differentially expressed between the mouse models and did not differentially change between L3 and Lc or between B6 and Bc were excluded.) We divided the mouse gene list into those corresponding to genes overexpressed in L3 relative to B6 tumours (192 genes), and those corresponding to genes overexpressed in B6 compared with L3 (187 genes). The statistical significance of enrichment between the mouse query genes and the human target signature was computed separately for the group up in L3 tumours and the group up in B6 tumours.

Defining tumour subtypes in human patients. We examined the expression of the 68 genes in the LOLES in 5 high-quality patient data sets in the Oncomine database[22,26–29], and found 19 genes to be upregulated in prostate cancer relative to benign prostate tissue in at least 3 of the 5 data sets (Supplementary Fig. S6o and Table S10). We used the LOLES to stratify the 79 primary tumour samples from the Glinsky data set[30] and the 131 primary tumour samples from the Taylor data set[22]. When multiple probes mapped to a single gene, the probe with the highest coefficient of variation was selected to represent the gene. To differentiate between aggressive and non-aggressive cases of prostate cancer, an event for the survival analysis was defined as a biochemical recurrence happening within five years. The LOLES classified these patients into high-risk and low-risk classes after k-means clustering with the kmeans function in R. The difference between these two classes was tested with respect to BCR survival time (Kaplan–Meier BCR-free survival curve) and the P value of this difference was computed with a log-rank test. For analysis of the Swedish cohort[25], we avoided overfitting by excluding patients used to construct the high-risk or low-risk human lethality signature (6 high-risk and 12 low-risk samples), resulting in 263 samples.

Pathway analysis. Enrichment of the mouse L3 versus B6 signature in human biological pathways was evaluated by GSEA using pathways collected in the REACTOME (ref. 52), KEGG (ref. 53) and BioCarta (http://www.biocarta.com/genes/allpathways.asp) databases. We mapped mouse genes to human orthologues using mouse–human orthologous relations from the Mouse Genome Informatics database (http://www.informatics.jax.org/), and selected those mapped genes that appeared in at least one biological pathway, resulting in 9,945 unique genes.

49. Banach-Petrosky, W. *et al.* Prolonged exposure to reduced levels of androgen accelerates prostate cancer progression in Nkx3.1; Pten mutant mice. *Cancer Res.* **67,** 9089–9096 (2007).
50. Hu, Y. & Smyth, G. K. ELDA: extreme limiting dilution analysis for comparing depleted and enriched populations in stem cell and other assays. *J. Immunol. Methods* **347,** 70–78 (2009).
51. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102,** 15545–15550 (2005).
52. Croft, D. *et al.* Reactome: a database of reactions, pathways and biological processes. *Nucleic Acids Res.* **39,** D691–D697 (2011).
53. Ogata, H. *et al.* KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **27,** 29–34 (1999).

nature cell biology

DOI: 10.1038/ncb2697

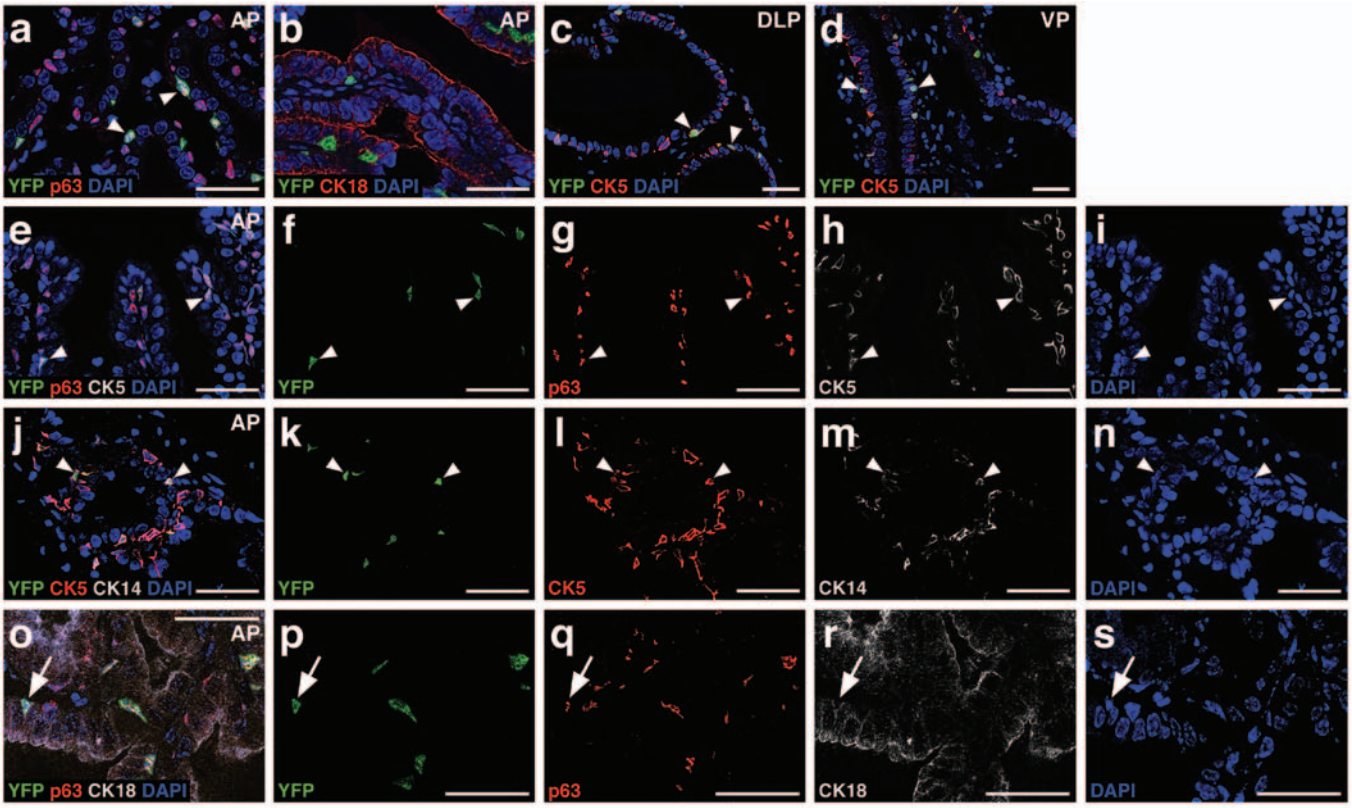

Figure S1 Cytokeratin 5 is a specific marker for basal and intermediate cells in all mouse prostate lobes. (**a**) Immunofluorescence staining showing co-localization of YFP with p63 in basal cells (arrowheads) of tamoxifen-induced *CK5-CreER^{T2}; R26R-YFP/+* anterior prostate (AP). (**b**) Lack of co-localization of YFP with the luminal marker CK18. (**c,d**) Co-localization of YFP with CK5 in the dorsolateral prostate (DLP, **c**) and ventral prostate (VP, **d**). (**e-i**) Overlay (**e**) and individual channels (**f-i**) demonstrating that all p63-positive cells (arrowheads) are also CK5-positive, and 99.1% of CK5-positive cells are also p63-positive (n=3,132/3,159, 3 animals analyzed). (**j-n**) Overlay (**j**) and individual channels (**k-n**) demonstrating that all CK14-positive cells (arrowheads) are CK5-positive, and vice versa. (**o-s**) Overlay (**o**) and individual channels (**p-s**) showing lineage-marking (YFP-positivity) of an intermediate cell (arrow) that is p63-positive and CK18-positive. Scale bars correspond to 50 microns.

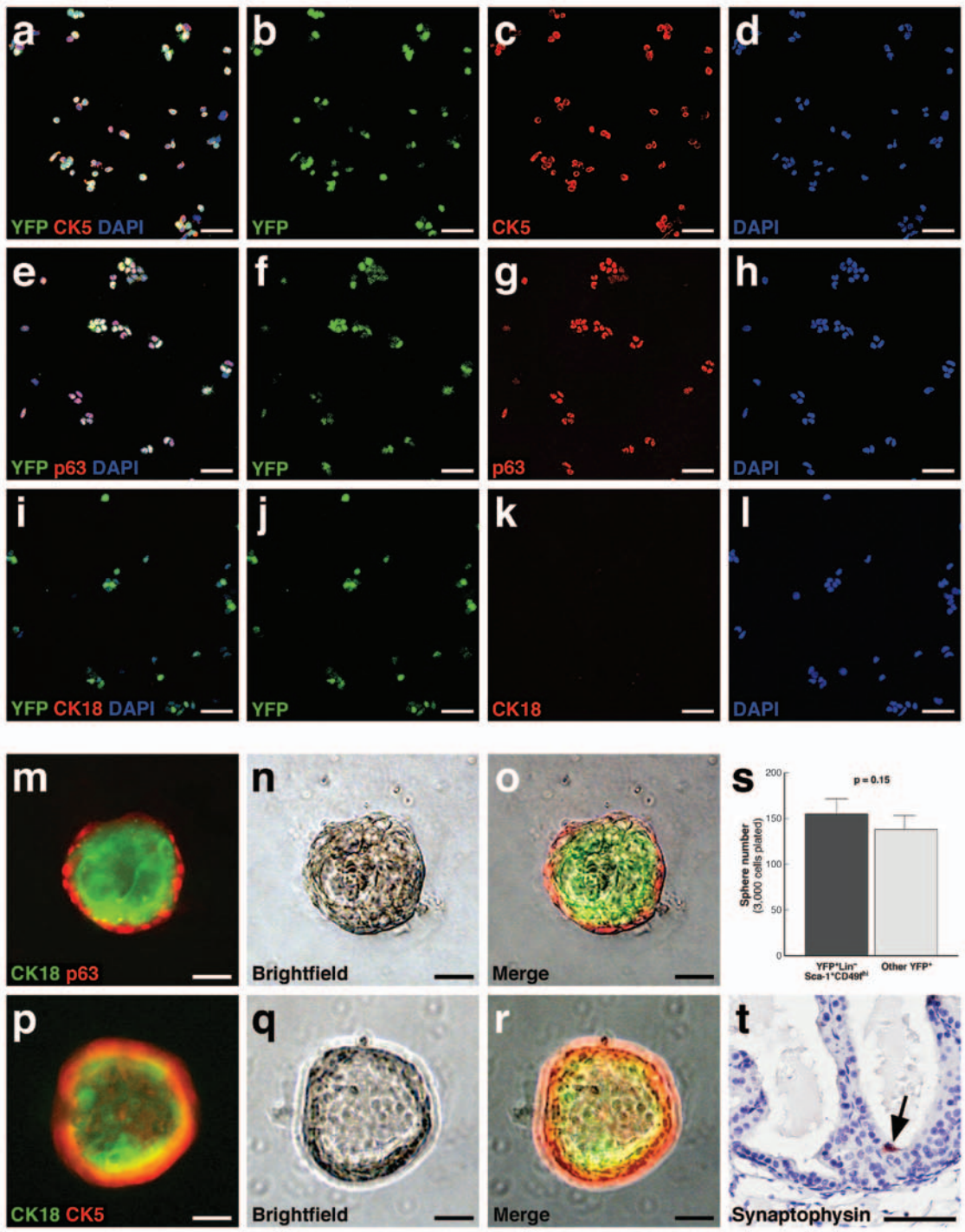

Figure S2 Marker analyses of basal cells isolated by flow cytometry and of the resulting spheres and renal grafts. (**a-l)** Purity of basal cells isolated by flow cytometry as shown in Fig. 1b, followed by immunofluorescence staining of cytospin preparations. (**a-d**) Overlay (**a**) and individual channels (**b-d**) showing co-localization of YFP and CK5 in the isolated basal cell population. (**e-h**) Overlay (**e**) and individual channels (**f-h**) showing co-localization of YFP and p63. (**i-l**) Overlay (**i**) and individual channels (**j-l**) showing absence of CK18 expression in the isolated basal cell population. (**m-r**) Spheres generated from basal cells after 7 days treatment with DHT, showing expression of luminal (CK18) and basal (p63, CK5) markers. (**s**) Quantitation of sphere formation from $YFP^{+}Lin^{-}Sca\text{-}1^{+}CD49f^{hi}$ cells (8.0% box in Fig. 1c), compared to the remaining YFP^{+} cells. 5 replicates from the same sorted population were performed for each group, and two sample t-tests were used for statistical analysis; error bars correspond to standard deviation and show variabilities beween replicates. (**t**) Detection of synaptophysin-positive neuroendocrine cell (arrow) in renal graft from basal cells. Scale bars in **a-r** correspond to 50 microns and in **t** to 100 microns; error bars in **s** correspond to one standard deviation.

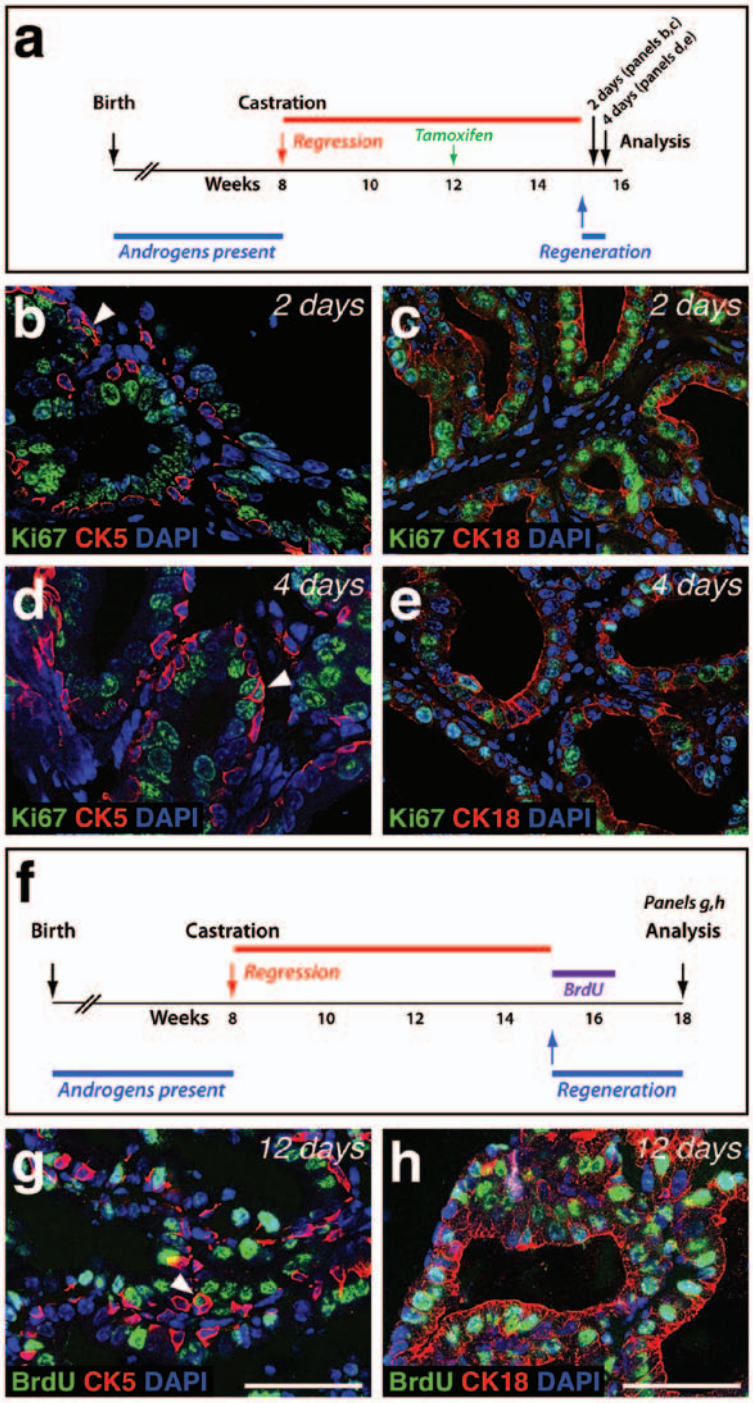

Figure S3 Analysis of cell proliferation during prostate regeneration. (**a**) Time course of prostate regression and regeneration in tamoxifen-treated *CK5-CreER^{T2}; R26R-YFP/+* mice. (**b-e**) Ki67 immunostaining to detect cell proliferation in CK5-positive basal cells (arrowheads) (**b,d**) and CK18-positive luminal cells (**c,e**) at 2 days after androgen administration to initiate regeneration (**b,c**) or 4 days after androgen administration (**d,e**). (**f**) *CK5-CreER^{T2}; R26R-YFP/+* mice were treated with BrdU during the first 12 days of prostate regeneration. (**g,h**) Immunostaining for BrdU incorporation to detect cell proliferation in CK5-positive basal cells (arrowhead, **g**) and CK18-positive luminal cells (**h**) at full regeneration. Scale bars in **b-e,g-h** correspond to 50 microns.

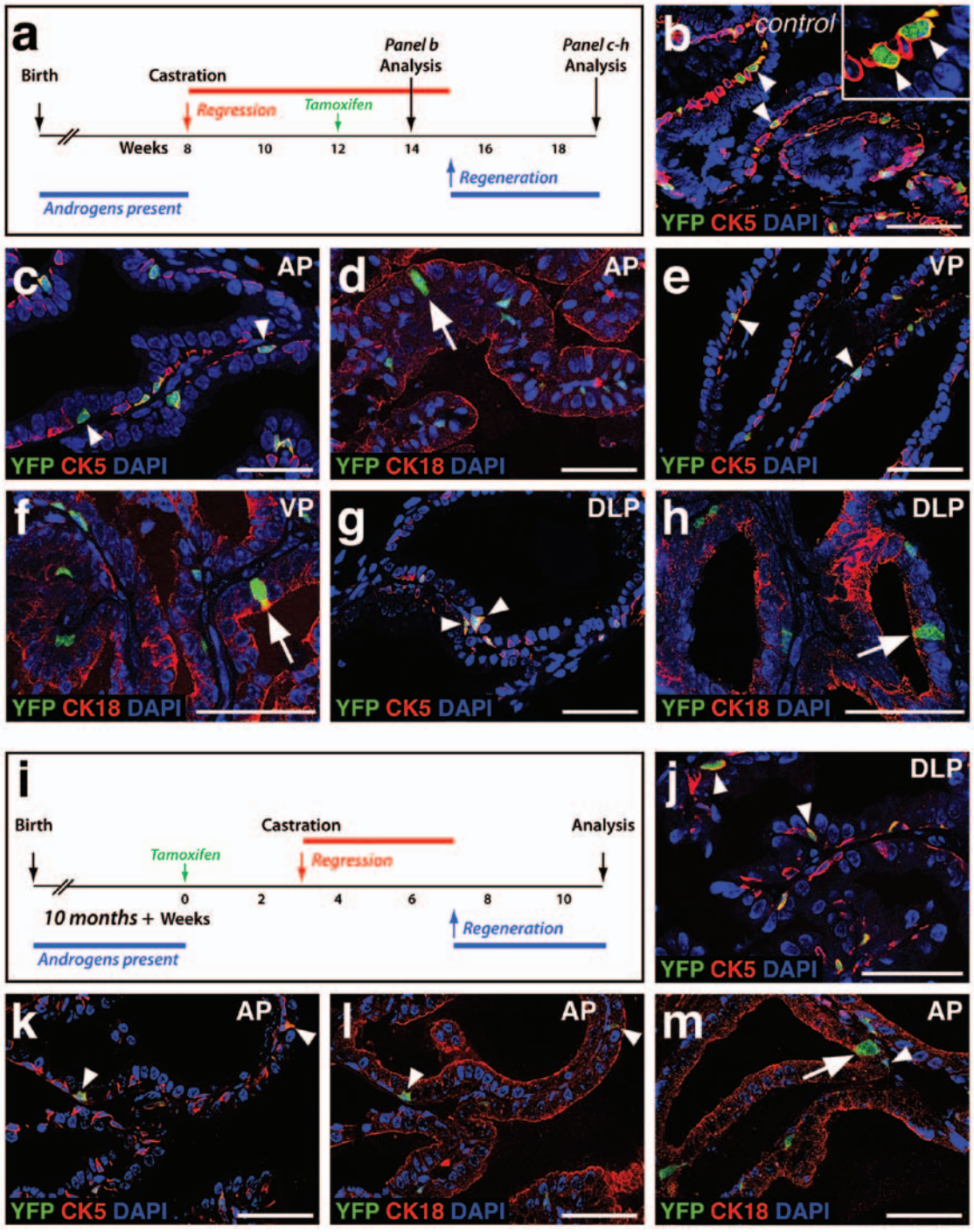

Figure S4 Additional data for lineage-tracing analysis of basal epithelial cells. (**a**) Alternative protocol for lineage-tracing during prostate regeneration, in which tamoxifen induction for marking of basal cells occurs in the regressed state. (**b**) Specific co-localization of YFP with CK5-expressing basal cells (arrowheads) after tamoxifen induction in androgen-deprived *CK5-CreER^{T2}; R26R-YFP/+* anterior prostate. (**c-h**) Most YFP^{+} cells express the basal marker CK5 (arrowheads, **c,e,g**) and not the luminal marker CK18 (panels d,f,h) after regeneration in the anterior prostate (AP, **c,d**), ventral prostate (VP, **e,f**), and dorsolateral prostate (DLP, **g,h**), although rare YFP^{+}CK18^{+} luminal cells (arrows, **d,f,h**) can be detected in all three lobes. (**i-m**) Lineage-tracing of basal cells during regeneration in older mice. (**i**) Strategy for lineage-tracing during prostate regeneration in mice starting at 10 months of age. (**j,k**) Co-localization of YFP with CK5-expressing basal cells (arrowheads) after regeneration in *CK5-CreER^{T2}; R26R-YFP/+* DLP (**j**) or AP (**k**). (**l**) General absence of YFP and CK18 co-localization in anterior prostate. (**m**) Detection of a rare YFP-expressing CK18-positive luminal cell (arrow) in anterior prostate. Scale bars correspond to 50 microns.

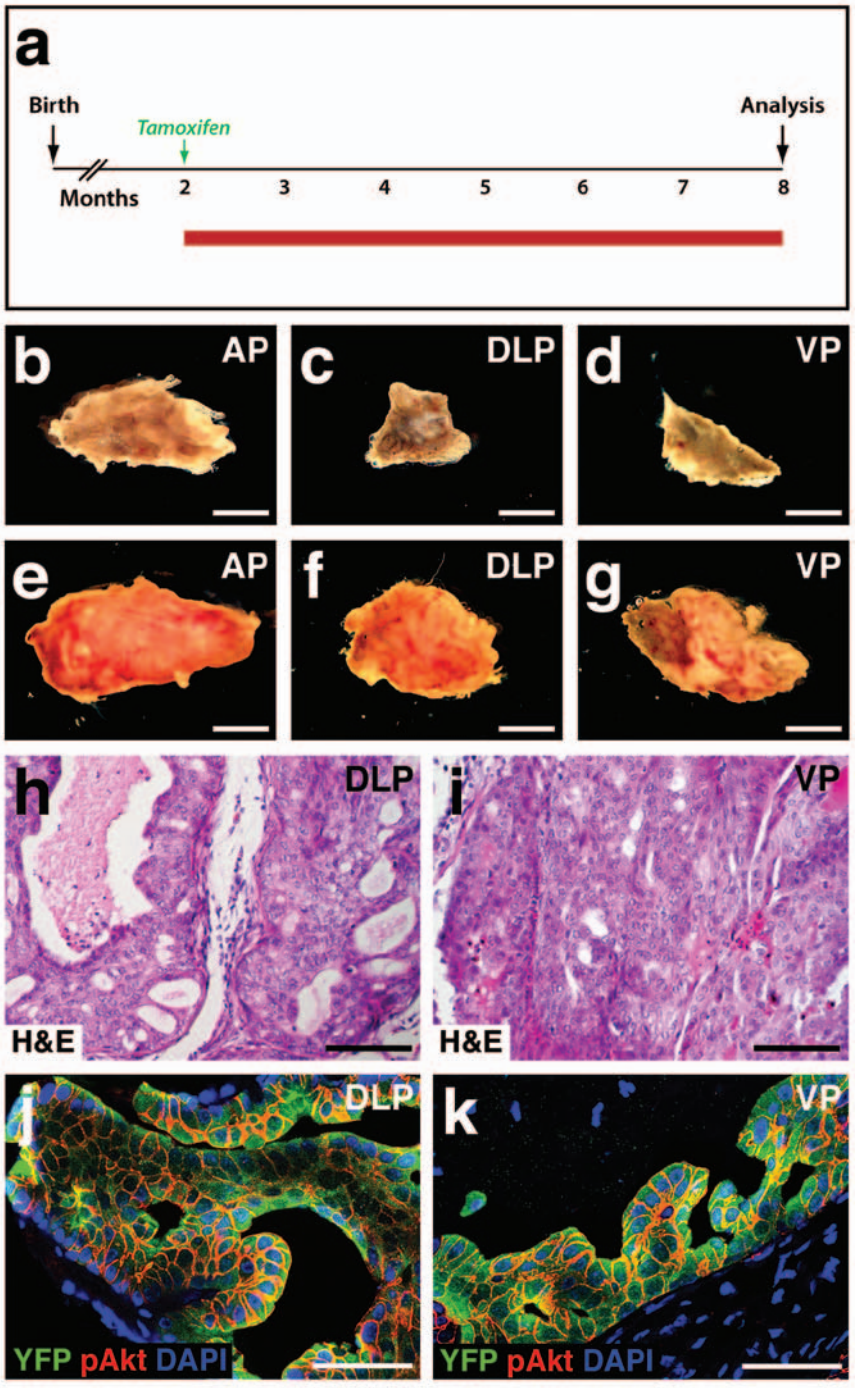

Figure S5 Phenotypes of PIN/tumor lesions following *Pten* deletion in basal cells. (**a**) Strategy for analysis of PIN/tumor lesion formation at six months after tamoxifen induction of hormonally-intact *CK5-CreER^{T2}; Pten$^{flox/flox}$; R26R-YFP/+* mice. (**b-g**) Whole-mounts of dissected anterior prostate (AP, **b,e**), dorsolateral prostate (DLP, **c,f**), and ventral prostate (VP, **d,g**) lobes from control uninduced *CK5-CreER^{T2}; Pten$^{flox/flox}$; R26R-YFP/+* mice (**b-d**) or tamoxifen induced mice of the same genotype (**e-g**). (**h,i**) Hematoxylin-eosin staining of PIN/tumor lesions in the DLP (**h**) and VP (**i**). (**j,k**) Widespread phosphorylated Akt (pAkt) expression at 6 months after induction in the DLP (**j**) and VP (**k**). Scale bars in **b-g** correspond to 2 mm, in **h,i** to 100 microns, and in **j,k** to 50 microns.

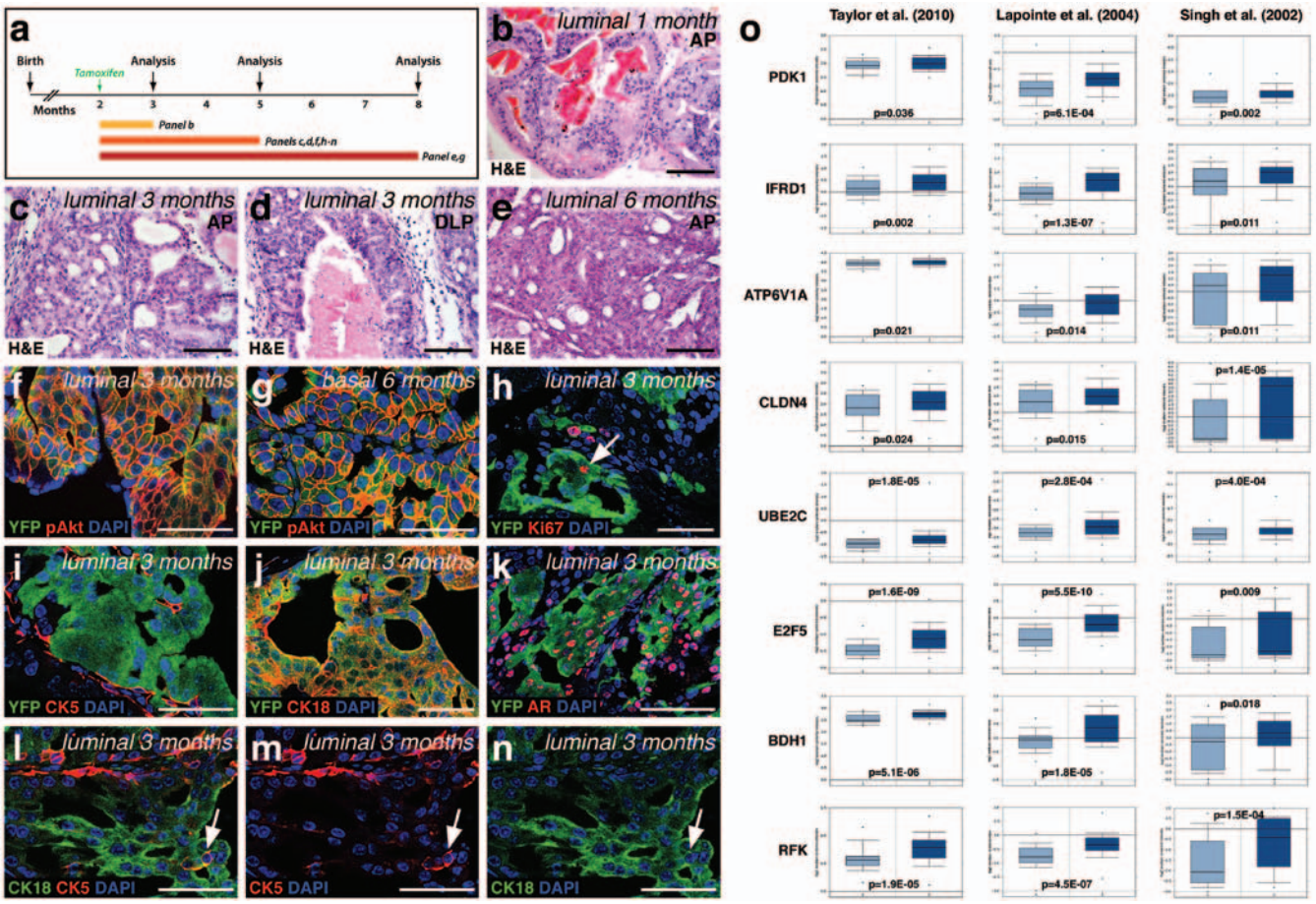

Figure S6 Phenotypes of PIN/tumor lesions following *Pten* deletion in luminal cells. (**a**) Time course for tumor formation after tamoxifen induction of hormonally-intact *Nkx3.1*$^{CreERT2/+}$*; Pten*$^{flox/flox}$*; R26R-YFP/+* mice. (**b-d**) Hematoxylin-eosin staining of anterior prostate (AP, **b,c,e**) and dorsolateral prostate (DLP, **d**) showing low-grade PIN at 1 month after induction (**b**), high-grade PIN at 3 months after induction (**c,d**), and high-grade PIN/carcinoma in situ at 6 months after induction (**e**). (**f,g**) Widespread phosphorylated Akt (pAkt) expression after three months of induction of *Nkx3.1*$^{CreERT2/+}$*; Pten*$^{flox/flox}$*; R26R-YFP/+* mice (**f**) is comparable to the pAkt immunostaining observed after six months of induction of *CK5-CreER*^{T2}*; Pten*$^{flox/flox}$*; R26R-YFP/+* mice (basal origin, **g**) **(h-n)** Marker analysis at three months after induction of *Nkx3.1*$^{CreERT2/+}$*; Pten*$^{flox/flox}$*; R26R-YFP/+* mice shows Ki67 immunoreactivity (arrow, **h**), lack of CK5 expression (**i**), expression of CK18 (**j**) and AR (**k**), and existence of CK5^{+}CK18^{+} intermediate cells (**l-n**). (**o**) Expression of selected genes of the luminal origin leading-edge signature in three different human patient datasets[22, 26, 28] in the Oncomine database, showing their up-regulation in human prostate cancer (dark blue boxplots) compared with benign prostate (light blue boxplots). Scale bars in **b-e** correspond to 100 microns, and in **f-n** to 50 microns.

Supplementary Tables

Table S1 Supporting data for quantitation of basal cell properties.

Table S2 Summary of histopathological phenotypes in mice with PIN/tumor lesions of basal or luminal origin.

Table S3 List of 200 differentially expressed genes between Basal 3 (3 months after induction) and Basal control (uninduced).

Table S4 List of 200 differentially expressed genes between Luminal 1 (1 month after induction) and Luminal control (uninduced).

Table S5 List of 200 differentially expressed genes between Basal 6 (6 months after induction) and Basal 3 (3 months after induction).

Table S6 List of 200 differentially expressed genes between Luminal 3 (3 months after induction) and Luminal 1 (1 month after induction).

Table S7 List of 200 differentially expressed genes between Luminal 3 (3 months after induction) and Basal 6 (6 months after induction).

Table S8 List of 200 differentially expressed genes between High-risk (death within 12 months) and Low-risk (survival for more than 192 months) human patient samples.

Table S9 List of biological pathways enriched in the Luminal 3 vs Basal 6 signature.

Table S10 List of genes in the leading edge from comparison of mouse Luminal 3 vs. Basal 6 signature with human high vs. low risk lethality signature.

Table S11 Antibodies used in this study.