Supplementary figure legends

Figure S1. Syntenic regions between human chromosome 21q and mouse chromosome 8q.

Almost all annotated protein-coding genes in the interstitial regions between *TMPRSS2* and *ERG* are well conserved between human (red) and mouse (blue).

Figure S2. IHC staining of prostate lesions in *Pb-Cre;T-ERG;Pten*$^{L/L}$ and *Pb-Cre;T-3Mb-Erg;Pten*$^{L/L}$ mouse models.

(**A**) IHC staining showing loss of SMA and p63 expression in moderately differentiated adenocarcinomas (with microducts) in *Pb-Cre;T-3Mb-Erg;Pten*$^{L/L}$ male mice. Scale bars are 100μm.

(**B**) IHC staining showing robust ERG expression (brown cells) in HG-PINs developed in both *Pb-Cre;T-ERG;Pten*$^{L/L}$ and *Pb-Cre;T-3Mb-Erg;Pten*$^{L/L}$ male mice, but not in HG-PINs developed in *Pb-Cre;Pten*$^{L/L}$ control males (ERG^{+} cells here are leukocytes and endothelial cells). Scale bars are 50μm.

Figure S3. Overview of mouse prostate tissues microdissected for genomic DNA PCR analysis and microarray expression profiling.

HG-PIN lesions were laser capture microdissected from both *Pb-Cre;T-3Mb-Erg;Pten*$^{L/L}$ (top) and *Pb-Cre;Pten*$^{L/L}$ control (bottom) prostates. In addition, poorly differentiated adenocarcinoma cells (tumor) were excised from *Pb-Cre;T-3Mb-Erg;Pten*$^{L/L}$ prostates. Genomic DNAs prepared from microdissected HG-PIN lesions and adenocarcinomas (tumor) from *Pb-Cre;T-3Mb-Erg;Pten*$^{L/L}$ prostates were used for PCR analysis to

determine Cre-mediated interstitial deletion. Total RNA isolated from microdissected tissues were used to perform gene expression analysis. Scale bars are 800μm.

Figure S4. Cre-mediated deletion of the interstitial region in *T-Δ-Erg/Pten*-null lesions.

(A) Schematic diagram showing PCR strategy (primers 1+2) to detect Cre-mediated excision of the 3Mb interstitial region and generation of the *Tmprss2-Erg* fusion at the same time.

(B) PCR analysis of genomic DNA prepared from HG-PIN lesions (mainly ERG^+) or large poorly differentiated adenocarcinomas (mainly ERG^-) isolated from *Pb-Cre;T-3Mb-Erg;Pten*$^{L/L}$ prostates by laser-capture microdissection confirmed Cre-mediated excision and creation of the *Tmprss2-Erg* gene fusion (detected by primers 1 and 2); primers specific for the wild type *Tmprss2* allele and the unexcised *T-3Mb-Erg* allele were used as controls. All three mice (Mouse #1-3) were *Pb-Cre;T-3Mb-Erg;Pten*$^{L/L}$ males.

Figure S5. Additional data showing the EMT phenotype in prostate lesions developed in *Pb-Cre;T-3Mb-Erg;Pten*$^{L/L}$ mice.

(**A-C**) GSEA results showing significant (FDR<0.25) enrichment of additional EMT gene set in *T-Δ-Erg/Pten*-null tumors in relation to HG-PIN lesions in *Pb-Cre;Pten*$^{L/L}$ control males (A), as well as highly significant enrichment of several EMT gene sets in *T-Δ-Erg/Pten*-null HG-PINs in relation to *Pb-Cre;Pten*$^{L/L}$ HG-PIN lesions (B) and in *T-Δ-Erg/Pten*-null tumors in relation to *T-Δ-Erg/Pten*-null HG-PINs (C). The gene sets are from the c2 CGP (chemical and genetic perturbations) collection of MSigDB (http://www.broadinstitute.org/gsea/msigdb/index.jsp).

(**D**) IF co-staining shows overlap of mesenchymal marker Vimentin with epithelial markers K8 and lineage marker YFP (genetic marking) in *T-Δ-Erg/Pten*-null tumors but not in HG-PIN lesions from the same mouse (although transcripts for EMT program have

already upregulated in such HG-PIN lesions in relation to HG-PIN lesions developed in *Pb-Cre;Pten*$^{L/L}$ control mice). Nuclei are counterstained with DAPI. Scale bars are 50μm.

(**E**) IHC staining for lineage marker YFP (based on *Pb-Cre*-mediated genetic marking) validated that tumor cells (yellow arrow) were derived from *Probasin* (Cre^+)-expressing prostate epithelial cells and not from stromal cells (blue arrow), which were YFP^-. Scale bar is 50μm.

Figure S6. Expression analysis of AR target genes in *T-Δ-Erg/Pten*-null and *Ras/Pten* prostate tumors.

(**A**) Expression of AR target gene *Fkbp5* and *Tmprss2* in *Ras/Pten* prostate tumors compared to *Pten*-null only prostate lesions. The data is based on GEO database accession # GSE34839.

(**B**) GSEA of *T-Δ-Erg/Pten*-null (left) or *Ras/Pten* (right) tumors in relation to *Pten*-null lesions for an androgen-responsive AR target gene set (based on Baena et al, PMID: 23512661). Note *Ras/Pten* tumors exhibit a higher level of downregulation of this AR target gene set (i.e., FDR=0.3, marginally significant, FDR<0.25 is considered as significant) than *T-Δ-Erg/Pten*-null tumors (FDR=0.79, not significant).

(**C**) Heat maps for GSEA in (B). The positions of *Fkbp5* in the heat maps are indicated. In the heat map, red, pink, light blue, and blue indicate high, moderate, low, and lowest expression levels of indicated genes.

Figure S7. Expression analysis of select ETS factor genes in prostate tissues.

(**A**) Relative expression values (normalized to that of *Erg*, =1) of several ETS factor genes, including *Ets2*, *Ets1*, *Erg* and *Etv1* in normal prostate tissues from various published microarray studies (GEO database accession numbers and microarray platforms are indicated). *Erg* and *Etv1* are involved in PCa-associated gene fusions and

are not expressed in normal prostate epithelial cells. All studies demonstrated a similar expression value of *Ets1* to those of *Erg* and *Etv1* (thus indicating that *Ets1* is also not expressed in normal prostates), whereas *Ets2* is highly expressed in normal prostate tissues.

(**B**) Relative expression values (normalized to that of *Erg* in *Pten*-null HG-PIN, =1) of several ETS factor genes showing expression of *Ets2*, but not *Ets1* in prostate lesions. Expression values are based on microarray data from this study.

Figure S8. Expression analysis of interstitial genes in *T*-Δ-*Erg/Pten*-null versus *Pten*-null HG-PIN lesions.

GSEA heat map showing slight downregulation of several interstitial genes (e.g., *BRWD1*, *ETS2*, *BACE2*) in *T*-Δ-*Erg/Pten*-null HG-PINs in relation to *Pten*-null HG-PINs. In the heat map, red, pink, light blue, and blue indicate high, moderate, low, and lowest expression levels of indicated genes. The processed expression data for this heat map is provided in Supplementary Table S1.