Genomic analysis of benign prostatic hyperplasia implicates cellular re-landscaping in disease pathogenesis

Lance W. Middleton, … , Robert B. West, Jonathan R. Pollack


Graphical abstract

Find the latest version:

http://jci.me/129749/pdf
Genomic analysis of benign prostatic hyperplasia implicates cellular re-landscaping in disease pathogenesis

Lance W. Middleton\textsuperscript{1}, Zhewei Shen\textsuperscript{1}, Sushama Varma\textsuperscript{1}, Anna S. Pollack\textsuperscript{1}, Xue Gong\textsuperscript{1, 2}, Shirley Zhu\textsuperscript{1}, Chunfang Zhu\textsuperscript{1}, Joseph W. Foley\textsuperscript{1}, Sujay Vennam\textsuperscript{1}, Robert T. Sweeney\textsuperscript{1}, Karen Tu\textsuperscript{1}, Jewison Biscocho\textsuperscript{1}, Okyaz Eminaga\textsuperscript{2}, Rosalie Nolley\textsuperscript{2}, Robert Tibshirani\textsuperscript{3, 4}, James D. Brooks\textsuperscript{2, *}, Robert B. West\textsuperscript{1, *}, Jonathan R. Pollack\textsuperscript{1, *}

Departments of \textsuperscript{1}Pathology, \textsuperscript{2}Urology, \textsuperscript{3}Biomedical Data Science, and \textsuperscript{4}Statistics, Stanford University School of Medicine, 300 Pasteur Drive, Stanford, California, 94305

*Address correspondence to James D. Brooks, M.D., Department of Urology, Stanford University School of Medicine, 300 Pasteur Dr, S287, Stanford, CA, 94305-5118, jdbrooks@stanford.edu; Robert B. West, M.D., Ph.D., Department of Pathology, Stanford University School of Medicine, 300 Pasteur Dr, L235, Stanford, CA, 94305-5324, rbwest@stanford.edu; Jonathan R. Pollack, M.D., Ph.D., Department of Pathology, Stanford University School of Medicine, 269 Campus Dr, CCSR-3245A, Stanford, CA, 94305-5176, pollack1@stanford.edu.

Conflict of interest statement: The authors have declared that no conflict of interest exists.
ABSTRACT

Benign prostatic hyperplasia (BPH) is the most common cause of lower urinary tract symptoms in men. Current treatments target prostate physiology rather than BPH pathophysiology and are only partially effective. Here, we applied next-generation sequencing to gain new insight into BPH. By RNAseq, we uncovered transcriptional heterogeneity among BPH cases, where a 65-gene BPH stromal signature correlated with symptom severity. Stromal signaling molecules BMP5 and CXCL13 were enriched in BPH while estrogen regulated pathways were depleted. Notably, BMP5 addition to cultured prostatic myofibroblasts altered their expression profile towards a BPH profile that included the BPH stromal signature. RNAseq also suggested an altered cellular milieu in BPH, which we verified by immunohistochemistry and single-cell RNAseq. In particular, BPH tissues exhibited enrichment of myofibroblast subsets, whilst depletion of neuroendocrine cells and an estrogen receptor (ESR1)-positive fibroblast cell type residing near epithelium. By whole-exome sequencing, we uncovered somatic single-nucleotide variants (SNVs) in BPH, of uncertain pathogenic significance but indicative of clonal cell expansions. Thus, genomic characterization of BPH has identified a clinically-relevant stromal signature and new candidate disease pathways (including a likely role for BMP5 signaling), and reveals BPH to be not merely a hyperplasia, but rather a fundamental re-landscaping of cell types.
INTRODUCTION

Benign Prostatic Hyperplasia (BPH) is the most common cause of Lower Urinary Tract Symptoms (LUTS) in men due to bladder outlet obstruction (1). Symptoms range from increased urinary hesitancy, urgency and frequency, to acute urinary retention. BPH is also associated with serious consequences including urinary bleeding, infections, bladder stones and renal failure. Histologic BPH is found in approximately 50% of men aged 50, and its prevalence increases about 10% each subsequent decade (2). Likewise, significant LUTS are documented to occur in about 10-20% of men aged 50-59 and increases to 1/3 of men by age 70-79. In parallel, prostate volume increases between age 40 to age 79, with the greatest increases appearing in the 6th and 7th decades of life (2-4).

BPH develops in the prostate transition zone and peri-urethral glands, and in advanced disease often appears grossly nodular (5). BPH represents a benign proliferative process and is not considered a precursor to prostate cancer. However, whether the prostatic hyperplasia reflects clonal cell expansions has never been examined. In BPH, increased cell numbers are found in both the epithelial and stromal compartments, though typically more so in the stroma (6, 7). Tissue architecture can vary, where in some cases glandular epithelial elements predominate, while in others fibromuscular-rich stroma comprises much of the BPH nodule, although most cases show a mixture of both histologies.

The pathophysiology of BPH is poorly understood, though androgen signaling, reactive stroma, and inflammation have been implicated (8). During prostate development, androgens appear to act primarily through the prostatic stroma, which in turn directs epithelial proliferation and differentiation (9). These stromal signals have been hypothesized to be re-activated during
Aging in BPH pathogenesis (5). Many stroma-derived growth factors appear to be controlled by androgen signaling in prostate tissue models (10-13) where they also act on prostatic stromal mesenchyme to drive proliferation and myofibroblastic differentiation (12, 14). Inflammation of the aging prostate has also been proposed to be an important driver of BPH (15-18). These inflammatory changes are thought to be linked to metabolic syndrome (19, 20) and may first exert their effects at the level of the prostatic stroma (17, 21-23).

Mirroring our incomplete understanding, BPH treatments have largely been developed empirically, and likely do not target the underlying disease mechanisms but rather the physiology of the entire prostate. Androgen inhibition (most commonly by 5-α-reductase inhibition) reduces prostate volume, but only partially relieves BPH symptoms mainly in men with significant prostate enlargement (24). Likewise, α-adrenoceptor blockade relaxes prostatic smooth muscle tone, but not all patients respond and α-blockade does not reduce the risk of needing surgery for BPH symptoms (25). Annual BPH treatment costs run in the billions of dollars (26). An improved understanding of BPH pathogenesis may suggest new strategies for precision therapy.

Genomics technologies have led to deeper understanding of human cancer, and more than 15 years ago researchers used DNA microarrays to begin to explore BPH (27, 28). Here, we revisit BPH using more sophisticated sequencing-based genomic approaches, including RNAseq to profile BPH transcriptomes and whole-exome sequencing to evaluate mutations and clonality. We discover new signaling molecules elevated in BPH tissue, including BMP5 and CXCL13, and a stromal transcriptional signature correlated with BPH symptoms. We also uncover significant changes in the cell types residing in both the BPH epithelium and stroma, reframing BPH and the disease process from a simple hyperplasia to a more fundamental re-landscaping of cells and tissue, with implications for disease prevention and treatment.
RESULTS

Transcriptome profiling identifies a stromal signature correlated with BPH symptoms

To investigate the transcriptional landscape of BPH, we performed RNAseq on BPH and matched normal prostate from FFPE tissue blocks, optimal for verifying histopathology. Cases were selected from radical prostatectomy specimens (performed for prostate cancer) with concurrent BPH, for which we had detailed assessment of urinary symptoms from pre-operative International Prostate Symptom Score (IPSS) with Bother Score (29). We selected cases from patients not on 5-α-reductase inhibitors or α-adrenoceptor blocking agents to eliminate confounding effects of BPH treatment. For about half the cases, we profiled matched histologically normal prostate tissue. For a smaller subset, we also profiled BPH stroma-rich nodules (Supplemental Figure 1), a recognized pathological entity characterized as discrete nodules of stromal cells with an absence of epithelial cells, to understand this BPH tissue variant and to aid in the interpretation of BPH signatures. In all, 65 samples passed rigorous RNAseq quality control metrics (see Methods), including 37 BPH samples, 19 normal prostate tissues, and 9 BPH stromal nodules (Supplemental Table 1).

Unsupervised hierarchical clustering (of normalized log₂ transcript levels) distinguished BPH from normal prostate, and the stromal nodules clustered distinctly from the BPH samples (Figure 1A and Supplemental Table 2). Gene expression features could be distinguished and interpreted based on characteristic marker genes, including features representing secretory epithelium/AR signaling (e.g., KLK3, TMPRSS2, KRT8), stroma (e.g. FBLN2, LUM, VCAN) (enriched in both BPH and stromal nodules), T cells/macrophages (e.g. CD2, CD5, LYZ) (enriched in BPH stromal nodules), and B cells (e.g., PAX5, CD79A, CXCR5) (also enriched in
BPH stromal nodules). Laser capture microdissection (LCM) was used to generate pure BPH epithelia and matched stroma from 5 independent cases followed by Smart-3SEQ (an RNAseq method developed for small cell numbers (30)). The transcript profiles confirmed that genes assigned to epithelial or stromal compartments in the bulk transcriptome data were expressed in those compartments (Figure 1A, left panel).

Across the BPH samples, unsupervised clustering revealed transcriptional heterogeneity, with most BPH cases partitioning into two main clusters (Figure 1A and Supplemental Figure 2). This subclass distinction appeared to be driven in part by a stromal gene feature, present with or without the stromal nodule samples (Supplemental Figure 2). Notably, when we compared patients with low vs. high expression of a 65-gene stromal signature, comprising the core of the stromal gene feature (Figure 1A and Supplemental Table 3), patients with a more pronounced stromal signature reported higher Symptom Score ($P=0.07$) and Bother Score ($P=0.02$) (Figure 1B). In contrast, a similarly derived 57-gene AR/Secretory signature did not associate with either Symptom Score or Bother Score (Figure 1C), nor did a comparable AR signature from the literature (31) (Supplemental Figure 3A and Supplemental Table 3). Indeed, plotting the correlation of genes with BPH Symptom and Bother scores (Figure 1A, right panel), the peak correlations occurred within the core stromal feature. Though we lacked an independent BPH cohort to validate the association between stromal gene-expression and patient-reported symptoms, we could build a predictor of Bother Score from the 65 stromal genes that by leave-one-out cross validation reduced the null model mean squared error by about 70% (Supplemental Figure 3B).

Transcriptome profiling uncovers signaling molecules including BMP5 upregulated in BPH
To more directly identify transcriptional differences between BPH and normal prostate, we performed a supervised analysis of the RNAseq data (SAM; (32)) (Figure 2A, B, and Supplemental Tables 4 and 5). In all, 301 genes were significantly more highly expressed in BPH, while 316 genes were more highly expressed in normal prostate (False Discovery Rate (FDR) <0.05). The genes upregulated in BPH included growth factors previously linked to prostatic hyperplasia, including fibroblast growth factors (FGF7), insulin growth factors (IGF2), and transforming growth factor beta (TGFB3) (8). However, topping the list were two seldom discussed signaling molecules, both markedly overexpressed in BPH: Bone Morphogenetic Protein 5 (BMP5; 50-fold elevated) and C-X-C Motif Chemokine Ligand 13 (CXCL13; 60-fold elevated). Elevated expression of both BMP5 and CXCL13 in BPH was confirmed by q-RT-PCR (Figure 2C). As yet, we have been unable to verify elevated BMP5 and CXCL13 at the protein level due to inadequate performance of antibodies in immunohistochemistry (IHC) and western blot.

Bone morphogenetic proteins are known to regulate diverse tissue morphogenetic processes in mammals (33). To investigate a mechanistic connection between BMP5 and BPH, we exposed immortalized prostate epithelial cells (RWPE-1) (34) to a range of physiologic concentrations of BMP5 protein (35). Canonical BMP receptor-SMAD pathway activation was evidenced by increased phospho-SMAD1/5/9 (Figure 3A) and induction of canonical BMP transcriptional targets (e.g., ID3, SMAD6) (36) (Supplemental Figure 4A). BMP5 addition led to a significant and reproducible 20% increase in RWPE-1 cell proliferation (Figure 3B). In addition, BMP5 treatment led to a more dispersed cell growth pattern, and to 30% increased cell migration by Boyden chamber assay (Figure 3C). In contrast, BMP5 addition to a different immortalized prostate epithelial cell line (BPH-1) (37) led to 15% decreased cell proliferation.
with no obvious cell dispersion (Supplemental Figure 4B). BMP5 also modestly reduced proliferation of WPMY-1 cells (Supplemental Figure 4B), an immortalized prostate myofibroblast line derived from the same prostate as RWPE-1 epithelial cells (38).

To further investigate the effects of BMP5 on cultured prostate cells, we profiled resultant changes in gene expression at 72hrs by RNAseq. Notably, BMP5 treatment of RWPE-1 prostate epithelial cells drove gene expression changes that recapitulated the changes seen in BPH compared to normal prostate. Specifically, by Gene Set Enrichment Analysis (GSEA) (39), the genes upregulated by BMP5 addition to RWPE-1 cells were significantly enriched for those genes more highly expressed in BPH (by our SAM analysis) ($P<0.0001$) (Figure 3D and Supplemental Table 6). Likewise, BMP5 addition to WPMY-1 prostate myofibroblast cells led to the induction of genes more highly expressed in BPH ($P<0.0001$), as well as genes of the BPH stromal signature ($P<0.0001$) (Figure 3E and Supplemental Table 6). As a specificity control, we confirmed that neither the RWPE-1 nor WPMY-1 transcriptional response to BMP5 showed enrichment of AR/secretory genes (Figure 3C, D). From the RNAseq data, RWPE-1 and WPMY-1 cells (as well as bulk BPH tissue) express presumptive BMP5 receptors (40), including type I receptors (BMPR1A, BMPR1B, ACVR1) and type II receptors (BMPR2, ACVR2A, ACVR2B) (Supplemental Table 7). RWPE-1 and WPMY-1 cells also express select other BMP ligands, though none are significantly upregulated in BPH (Supplemental Table 7). Our finding that BMP5 addition in two different cell lines mimics the transcriptional changes of BPH suggests a role for BMP5 in driving BPH pathogenesis.

We also investigated CXCL13, the other ligand highly expressed in BPH tissue. From our RNAseq data, neither RWPE-1 nor WPMY-1 cells express the only known CXCL13 receptor, CXCR5, which we confirmed by qRT-PCR (Supplemental Table 8). Moreover,
addition of recombinant CXCL13 to RWPE-1 and WPMY-1 cells did not impact cell proliferation or migration (Supplemental Figure 4C, D). Rather, CXCL13 has been characterized primarily as a B-cell chemoattractant (41). Indeed, from our BPH transcriptional heatmap (Figure 1A), we had identified a B-cell gene feature that includes the CXCR5 gene, is prominent in BPH stromal nodules, and is correlated with stromal CXCL13 expression (Supplemental Figure 4E). Nonetheless, in contrast to CXCL13, the B cell feature did not show a significant increase in BPH vs. normal prostate tissue.

Transcriptional changes uncover cell subsets altered in BPH

In analyzing the genes with significantly decreased expression in BPH (vs. normal prostate), several stood out as markers of specific cell types. For example, chromogranin A (CHGA), a marker of neuroendocrine (NE) cell lineage, was downregulated in BPH (Figure 2B). To further investigate this finding, we constructed a Tissue Microarray (TMA) containing over 100 normal prostate and BPH tissues, including most of the cases used for RNAseq. Immunohistochemistry (IHC) demonstrated significant depletion of a subpopulation of CHGA-expressing neuroendocrine cells in the epithelium of BPH compared to matched normal (Figure 4A-I).

Estrogen Receptor (ESR1) transcript was also significantly decreased in the BPH samples (Figure 2B). Estradiol and ER signaling have been implicated in rodent and canine studies as effectors of BPH (42, 43). By IHC, we observed that in normal prostate ESR1 was expressed in a distinctive subset of periglandular fibroblast-like stroma cells, while in corresponding BPH stroma ESR1 protein was not detected (Figure 4J-N). Consistent with this finding, Ingenuity Pathway Analysis (IPA) of the transcriptome data identified estradiol as by far ($P=10^{-18}$) the top-ranked predicted regulator diminished in BPH (Supplemental Table 9). Based on these findings,
BPH stroma appears to have lost a subset of ESR1+ fibroblast-like cells, with concomitant loss of downstream estrogen receptor signaling.

To investigate whether additional cell lineages might be affected in BPH, we carried out additional IHC using known lineage markers. We observed a trend towards fewer AR-positive nuclei in BPH epithelium, significantly reduced desmin (DES) staining (myofibroblasts/smooth muscle cells) in BPH stroma and BPH stromal nodules, a trend towards fewer CD3+ cells (T cells) in BPH, significantly increased CD20+ cells (B cells) in BPH stromal nodules, and significantly increased CD163+ cells (macrophages) in BPH and BPH stromal nodules (Supplemental Figure 5). Vimentin (VIM) staining (fibroblasts, myofibroblasts) showed little variation among BPH samples by our semi-quantitative IHC assay, and so was not scored.

To independently evaluate altered cell subsets in BPH, we carried out single-cell RNAseq (scRNAseq) of matched BPH and normal tissue from a single prostate. Fresh tissue was enzymatically disaggregated, and then single cell transcriptomes sequenced using the 10X Genomics Chromium workflow. In total, we analyzed 9,373 single cells with a median 1,891 genes expressed per cell. Dimensional reduction and clustering revealed distinct cell clusters (identifiable by characteristic marker genes) representing epithelial cells (basal, intermediate and secretory), myofibroblasts, immune cells and endothelial cells (Figure 5A-D and Supplemental Figure 6A-D). The BPH tissue showed depletion of neuroendocrine cells (CHGA; Figure 5E-F, I-J), as well as enrichment of a myofibroblast subset expressing BMP5 and CXCL13 (Figure 5G-H, K-L); both findings were consistent with the bulk RNAseq data. In contrast, ESR1 transcript was detected in a subset of myofibroblasts from both the normal prostate and BPH samples (Supplemental Figure 6E-H). In the single prostate sampled, the BPH epithelial fraction also exhibited a shift towards more mature secretory epithelial cells (Figure 5A, E, I).
Exome sequencing reveals expansion of specific cell clones in BPH

To address whether the underlying hyperplasia of BPH might represent a clonal process, we performed whole exome sequencing (WES) on BPH tissues from 18 patients (fresh or FFPE tissue) along with matched normal DNA from either normal prostate or lymph node. Analysis of the WES data revealed 38 “high-evidence” (see Methods) SNVs acquired among the BPH tissues (Figure 6A, B and Supplemental Table 10), with a median of 2 SNVs per patient (range, 0 to 6). We successfully validated 2/2 (100%) SNVs by PCR/Sanger sequencing, and 15/16 (94%; randomly selected) by PCR amplicon deep-resequencing (Figure 6C, D and Supplemental Table 11). The SNV allelic frequencies (many as high as 20-30%) suggest that these presumed heterozygous mutations were present in up to 50% of the cells in the sample. The finding of an enriched population of cells with a somatically acquired mutation implies a clonal outgrowth of a population of cells from a single cell.

The 38 high-evidence SNVs occurred across 33 genes, of which 2 genes were recurrently mutated; dentin sialophosphoprotein (DSPP) harbored 5 SNVs among 3 different BPH samples (with two adjacent nucleotides identically mutated in 2 different BPH samples), and titin (TTN) was represented by 2 SNVs over 2 different BPH samples. The 38 SNVs included frameshifting deletions, stop-gain mutations, and non-synonymous substitutions, many predicted (by evolutionary conservation) to be deleterious. Interestingly, the 33 affected genes were significantly enriched for Cosmic Cancer Gene Census (44) tier-1 genes (enriched genes: BARD1, CREBBP, NUP98, ZNF384) (P=0.01; Hypergeometric test). There was no enrichment for genes differentially expressed in BPH vs. normal prostate (from our SAM analysis).
Recent studies have shown that tissues can acquire SNVs during the cell expansion and turnover with normal tissue development and maintenance (45). To further evaluate the BPH acquired SNVs, we performed a reverse analysis seeking to identify acquired SNVs in the matched normal tissue (using the BPH tissues as reference). Notably, the proportion of high-evidence SNVs identified among the BPH samples (forward analysis) was 7-times higher than among the normal tissues (reverse analysis) ($P<0.0001$; chi-squared test), suggesting the BPH acquired SNVs are linked to the disease process.
DISCUSSION

Our study describes a first sequencing-based genomic landscape of BPH. We observed transcriptional heterogeneity among BPH samples, including varied expression of a BPH stromal gene feature. Notably, elevated expression of a 65-gene BPH stromal signature (the core of the stromal gene feature) was associated with worse patient-reported LUTS and greater bother from those symptoms. This finding highlights the importance of prostatic stroma in driving BPH symptoms, and agrees with previous observations that increased fibrosis and stiffness of the prostatic tissues is correlated with worse LUTS (46). Our findings also echo a prior microarray study that reported an “extracellular matrix” gene-expression feature that distinguished symptomatic from asymptomatic BPH (28). However, whether that feature corresponds to our stromal feature is unclear since those microarray data are not available for reanalysis. That our 65-gene stromal signature correlates with symptoms suggests that it is capturing an intrinsic tissue property robust to tissue sampling variables. Whether the signature reflects an increased stromal proportion or else increased stromal activation remains to be determined. In addition to validating the stromal signature in independent datasets, its association with treatment response to 5-α-reductase inhibitors and α-adrenoceptor blocking agents should be tested. We also observed that an androgen-signaling signature (at least in prostatic epithelia) does not correlate with LUTS, underscoring that androgen signaling may play a permissive role for prostate growth but may not be an underlying driver of BPH.

The genes we found more highly expressed in BPH, which include many of the stromal signature genes, represent new candidate drivers and therapeutic targets. For example, BMP5 and CXCL13 transcript levels were highly overexpressed (at least 50-fold), and could be
targetable as both are signaling molecules. BMP5 is a member of the bone morphogenetic protein family (and TGF-β superfamily) that has broad roles in tissue morphogenesis (33). In prior microarray studies, BMP5 was listed among genes found more highly expressed in the transition zone of the normal prostate (where BPH arises) (47) and in BPH (27), though it has never been investigated as a BPH driver. Buttressing our RNAseq results, we found that the addition of BMP5 protein to RWPE-1 prostate epithelial cells enhances cell proliferation. Although the proliferative effect is modest (20% increased cell numbers at 96hrs), the cumulative effect over years could more than explain the hyperplasia observed in BPH. BMP5 treatment of RWPE-1 cells also led to increased cell dispersion and transwell migration, features of epithelial-mesenchymal transition (EMT). These finding are consonant with a prior study that reported features of EMT in BPH tissue (48). However, since the increased cell proliferation and dispersion were not replicated in a second immortalized prostate epithelial line (BPH-1), caution is warranted in not over-interpreting the findings. Likewise, that BMP5 modestly reduced WPMY-1 cell proliferation, whilst increased stromal cells are observed in BPH, suggests that the net impact of BMP5 on prostate tissue myofibroblasts is modulated by other signaling molecules.

Most remarkably, BMP5 treatment of both prostate epithelial (RWPE-1) and prostate myofibroblast (WPMY-1) cells resulted in gene expression programs that mirrored those upregulated in BPH tissue, and (for WPMY-1 cells) in BPH stroma. Thus, BMP5 likely acts directly on prostatic epithelia and stroma to effect BPH changes, though possibly also indirectly by modulating stromal cells to secrete paracrine factors. Future work will need to confirm BMP5 overexpression at the protein level, define the most important receptors (in both prostatic epithelia and stroma) from candidates observed in our RNAseq data, and evaluate BMP5 signaling as a therapeutic target. Importantly, screening drugs/molecules for their ability to revert
BPH cells/tissues to more closely mimic normal prostate transcriptomes might provide a valuable approach to evaluate new candidate BPH therapies.

CXCL13, the other highly overexpressed signaling molecule, has not been studied in the pathogenesis of BPH. CXCL13 is a known as a B-cell chemoattractant (41), and B cell numbers were largest in BPH stromal nodules, where CXCL13 was most highly expressed. However, apart from stromal nodules, B cells (or the B-cell signature) were not more prevalent in BPH compared to normal tissue. And while a different chemokine (CXCL12; not found in our BPH transcript set) has been reported to enhance proliferation of prostatic epithelial cells in vitro (49), we did not observe CXCL13 effects on prostate cells. Therefore, the relevance of elevated CXCL13 to the BPH disease process (apart from BPH stromal nodules) remains uncertain, but warrants further studies.

Beyond BPH genes and signatures, our bulk transcriptome analysis implied changes in the resident cell types, which we confirmed and extended by IHC and single-cell RNAseq analysis. Our combined studies identified marked depletion of neuroendocrine cells in the BPH epithelial compartment. Neuroendocrine cells are a relatively uncommon cell type within the normal prostate epithelium, and produce neurosecretory products (including serotonin) with paracrine effects on prostatic epithelium (50). A previous IHC study also reported reduced neuroendocrine cells in BPH tissue (51), though this finding has been contested (52), and serotonin has been implicated in negatively regulating prostate growth (53). Equally interesting was our discovery of an ESR1+ stromal cell type (with fibroblast morphology) residing adjacent to epithelium, present in normal prostate but absent from BPH. Moreover, pathway analysis identified estradiol as the top regulator decreased in BPH compared to normal tissue. While speculative, these findings suggest that ER signaling in prostatic stroma might constrain BPH, in
marked contrast to rodent and canine models where estrogen administration produces prostatic enlargement (42, 43). Alternatively, the depletion of ESR1+ stromal cells might reflect the relative expansion of other stromal cell types, or else loss of ESR1 expression rather than cell type depletion (though raising the question of what defines cell type vs. cell state (54)).

Concomitant with the observed cell type depletions, our bulk RNAseq and single-cell RNAseq data suggested expansions of other cell types, including stromal (likely myofibroblast) subsets expressing BMP5 and CXCL13. Additional epithelial cell type shifts were also noted in the single scRNAseq sample, though not supported from our bulk RNAseq data; additional studies are warranted. Lastly, immunostaining with cell lineage markers supported additional cell type shifts in BPH, including alterations in myofibroblast/smooth muscle proportions (decreased desmin staining), and increased tissue macrophages.

Whether BPH nodules represent clonal cell expansions has not been studied. Remarkably, by exome sequencing we discovered somatically-acquired mutations in BPH. Many of these mutations are predicted to affect protein function, though it remains unclear whether they are growth promoting “drivers” or passenger genes that have no functional role in normal prostate biology. The genes did show enrichment for Cosmic Cancer Census Genes, suggesting the mutations might have a functional role in BPH development, though the gene numbers are small. We also found recurrently mutated genes across BPH samples, including TTN and DSPP which have been reported mutated in human cancers (55, 56). TTN functions in striated muscle elasticity (57), while DSPP functions in dentin mineralization (58), neither with an obvious connection to BPH. Moreover, TTN is a large late-replicating gene, a gene class that incurs increased mutations but is thought unlikely to represent cancer drivers (55), and DSPP contains a long stretch of triple codon repeats, and therefore may be subject to increased mutation or
sequence read-mapping errors. While further studies are needed to determine whether the mutations observed in any of the 33 mutated genes are functionally relevant to BPH, the somatic SNVs nonetheless serve as lineage markers and are indicative of clonal cell expansions. We found significantly more SNVs in BPH compared to normal tissue (adjacent prostate or lymph node), suggesting that SNVs were acquired in recent cell expansions related to the BPH disease process, rather than during organ development. Moreover, the alternate allele frequencies suggest that the clonal population represents a substantial portion of the cell population in the BPH tissues. Future studies will determine which cell type(s) are clonally expanded, and whether they are cellular drivers of the BPH disease process.

BPH has been thought of as prostate enlargement due to cellular hyperplasia. Taken together, our transcriptome and exome studies support a rather different picture, of a disease resulting from a fundamental re-landscaping of cell types. It is possible that cell type depletions represent passive losses due to the expansion of others. However, the known intricacies of epithelial-stromal crosstalk and tissue homeostasis suggest that these cell type shifts are more likely to be intimately connected to disease pathogenesis and may fundamentally underlie the pathobiology and resulting symptoms. Based on our genomic characterization, we propose that BPH should be studied not only as altered molecules and signaling pathways, but as a fundamental re-landscaping of cell types and tissue. Analysis of single cells in co-culture models and from disaggregated prostate tissues and in situ in normal and diseased tissues will be illuminating.
METHODS

Prostate tissue specimens
Tissue specimens were acquired from radical prostatectomies for the diagnosis of prostate cancer that had concurrent BPH. All experiments involving human subjects were approved by the Stanford Institutional Review Board (IRB), and patient consent was obtained pre-operatively by the surgeon, surgical staff or Stanford Tissue Bank staff. For RNAseq, cases were selected from prostatectomies performed between 2011 and 2013, for which patients had completed the International Prostate Symptom Score (IPSS) with Bother Score prior to surgery. Histopathologic samples of formalin-fixed paraffin-embedded tissues were reviewed by a pathologist (RBW) and 1mm diameter core punch biopsies were taken from the blocks to ensure that the histological entities were clearly identified and not contaminated by other histologic changes such as prostate cancer. Cores were harvested from regions of BPH (mixed epithelia/stroma), BPH pure stromal nodules, and normal peripheral zone of the prostate far away from any cancer and without other pathologic changes. For WES, cases were selected from freshly-frozen tissue and FFPE blocks, archived from 2005-2013.

RNAseq and RNAseq data analysis
RNA was isolated from FFPE cores using Qiagen RNeasy FFPE kit, and then quantified by NanoDrop spectrophotometer and Qubit fluorometer. Barcoded RNAseq libraries were constructed from 1μg input RNA using Illumina TruSeq Stranded Total RNA LT kits with Ribo-Zero Gold depletion of rRNA. Barcoded libraries were then quantified by NanoDrop, qualified
by Agilent BioAnalyzer, pooled (6 libraries per lane), and sequenced (unpaired, 100bp) on an Illumina HiSeq2000 instrument (Stanford Genome Sequencing Service Center) to an average depth of 39 million reads. Barcodes were deconvoluted using Illumina CASAVA software and reads quality-assessed using Babraham Institute FastQC software. Reads were then mapped to the RefSeq transcriptome using TopHat and Cufflinks (59), and transcripts quantified as Reads Per Kilobase of transcript per Million mapped reads (RPKMs). Not all of the FFPE RNAs yielded high-quality sequencing libraries; 65 samples (84%) passed rigorous quality-control criteria based on RPKM histograms (>30% of genes with RPKM≥1) and RPKM values for housekeeping genes (GAPDH>20; ACTB>100; B2M>40). RNAseq data were then further processed by eliminating poorly measured genes, as those absent from more than 50% of samples or having average expression <0.1 RPKM. Gene RPKM values were then bottom-thresholded at 0.1 and transformed to log2 space. RNAseq libraries had been constructed by two different individuals, necessitating batch normalization by separately median centering genes within each batch, and then re-centering the combined dataset; equivalent results were obtained by batch normalization using ComBat (60). Subsequent analyses of variably-expressed genes included 2,548 genes with standard deviations ≥1.2 across samples. Unsupervised hierarchical clustering (Person correlation; average linkage) was performed using Cluster 3.0 software (61) and visualized using Java Treeview (62). High vs. low expression of a stromal (65 genes) and AR/Secretory (57 genes) signature was defined using a median cutoff (across samples) of the average expression of all signature genes within each sample. In two patients with more than one BPH sample, signature scores were averaged prior to analysis. Gene-based models to predict BPH symptoms were built using lasso (63), implemented in R, and performance (added value) evaluated by the reduction of the mean squared error compared to the null model (no genes).
Supervised analysis of differentially expressed genes was done using Significance Analysis of Microarrays (SAM) (two class unpaired, t-statistic) (32), though DESeq2 (64) performed comparably. Pathway analyses were done using Ingenuity Pathway Analysis (IPA). qRT-PCR validation of BMP5 and CXCL13 was done using TaqMan Gene Expression Assays (Hs00234930_m1 and Hs00757930_m1) (Life Technologies) on an AB7500 Fast Instrument.

3SEQ of LCM samples
To confirm assignments of gene expression features to epithelial vs. stromal compartments, laser capture microdissection (LCM) was performed on 5 independent BPH cases, using an Arcturus XT LCM system, and transcriptomes profiled by SMART-3SEQ exactly as described (30). Genes significantly more highly expressed in BPH epithelia or stroma were identified by DESeq2 (64).

Prostate cell culture experiments
RWPE-1, WPMY-1 and BPH-1 cell lines were obtained directly from the ATCC and DSMZ cell culture repositories, cultured as directed, and passaged for no more than 3 months. Cells (25-40K) were plated in 6-well plate wells in duplicate, and then recombinant human BMP5 (R&D Systems), CXCL13 (R&D Systems), or vehicle control added daily with media replacement. Relative cell counts were determined at 96hrs by WST-1 assay (Sigma). All cell growth experiments were repeated at least 3 times with comparable results. Western blots were used to quantify canonical SMAD pathway activation (30 min post BMP5 treatment), using qualified primary antibodies (Supplemental Table 12), chemiluminescence detection, and ImageJ quantification. Boyden chamber (transwell) migration assays were carried out in triplicate, using
polyethylene terephthalate 8µM pore inserts (Falcon) and a chemotactic gradient of complete media against a 1:20 dilution into base media. After 48hrs, migrated cells were stained with crystal violet and counted. For transcriptome sequencing studies, RWPE-1 and WPMY-1 cells (assayed in duplicate) were plated, and 250ng/ml BMP5 (or vehicle control) added with daily media replacement for 72hrs. RNA was isolated using Qiagen RNeasy kit, RNAseq libraries generated using TruSeq Stranded Total RNA LT kit (WPMY-1) and Illumina TruSeq RNA Library Prep Kit v2 (RWPE-1), and barcoded RNAseq libraries sequenced and resultant data processed as above. RPKM values were then averaged for the duplicate platings, and enrichment for gene sets was evaluated by Gene Set Enrichment Analysis (GSEA) (39).

**Immunohistochemistry on Tissue Microarrays**

A tissue microarray was constructed using a manual Beecher instrument, and included duplicate 1 mm cores from over 100 BPH and matched normal prostate tissues. TMA 4µM sections were processed for antigen retrieval (Ventana or Leica proprietary solution), and then used for IHC with primary antibodies (listed in Supplemental Table 12), followed by HRP-conjugated secondary antibody and DAB- chromogenic detection (Stanford Pathology IPOX lab). IHC was interpreted by a trained pathologist (RBW) (scoring criteria listed in Supplemental Table 12), and scores for duplicate cores averaged.

**Single-cell RNAseq**

Fresh prostate tissue was acquired from radical prostatectomy cases, where BPH was identifiable grossly by transition zone nodules, and normal prostate tissue harvested from the peripheral zone opposite that having biopsy-confirmed cancer. Tissue was disaggregated using
collagenase/hyaluronidase, trypsin and dispase, following StemCell Technologies’ Prostate Tissue Dissociation protocol. Contaminating red blood cells were lysed using ammonium chloride solution (StemCell Technologies), and single cells filtered through a 40µM cell strainer. Half the remaining cells were further purified by flow cytometry, gating on viable (propidium iodide-negative) single cells. Single-cell RNAseq was then performed using the 10X Genomics Chromium workflow (by the Stanford Functional Genomics Facility), entailing microdroplet single cell capture (with cell and molecular barcodes), cDNA synthesis, pooling and library preparation, followed by sequencing on an Illumina HiSeq4000. Cells and transcripts were then enumerated using the 10X CellRanger pipeline and visualized using 10X Loupe software. Data from flow sorted and unsorted cells were indistinguishable and combined for the illustrations shown. Note, the tissue processing procedure appears to have enriched for epithelial over stromal cell populations.

**Whole-exome sequencing**

WES was carried out on individual BPH nodules, mostly comprising mixed epithelia and stroma. DNA was extracted using the Qiagen DNeasy kit (fresh tissue) or QIAamp DNA FFPE Tissue Kit, and quantified by Nanodrop and Qubit. Whole-exome sequencing was done using the Agilent SureSelect Human All Exon V5 kit, following the modifications recommended for FFPE-derived DNA samples. Barcoded WES libraries (4 per lane) were sequenced on a HiSeq2000 to an average 60X mean bp coverage. Raw reads were aligned to the hg19 human genome using BWA (65), and Picard was used to mark duplicates. The GATK 3.x best practices were followed for Indel Realignment and Base Recalibration. Callable regions were calculated using bedtools followed by an ensemble approach of four variant callers: MuTect, VarScan2,
VarDict, and Freebayes (66-69). Exonic and splicing SNVs called by at least two of the four
callers were further processed by ANNOVAR (70) and custom python scripts for variant
annotation, filtering, and ranking. “High-evidence” SNVs were defined as those with at least 10
alternate allele reads in the BPH sample, and 0 alternate allele reads in the matched normal
sample. Selected SNVs were validated by PCR/Sanger sequencing (PCR primers available in
Supplemental Table 13). Additional SNVs were validated by amplicon deep re-sequencing (PCR
primers available in Supplemental Table 13), with multiplexed barcoded amplicons sequenced
on an Illumina NextSeq.

Data availability
All RNAseq and WES data are available from dbGAP (accession phs001698.v1.p1). TMA IHC
images are available from TMAD (https://tma.im/cgi-bin/home.pl).

Statistics
Unless otherwise indicated, a two-sided Student’s t-test was performed to assess statistical
significance, using Prism version 7.0 (GraphPad), where $P < 0.05$ was considered significant.
Two-class analysis of RNAseq data was done using SAM (32), which estimates false discovery
rates against randomly permuted samples. Gene set enrichment analysis was done by GSEA
(39), with significance determined by Kolmogorov-Smirnov statistic.

Study approval
The study was approved by the Stanford University Institutional Review Board. Written
informed consent was received from participants prior to including in the study.
AUTHOR CONTRIBUTIONS

LWM, ZS, JDB, RBW and JRP conceived and planned the studies; LWM, ZS, ASP, SXV, SZ, CZ, KT, JB, OE and RN performed experiments; XG, JWF, SV, RTS, RT, JDB, RBW and JRP analyzed data; LWM, JDB, RBW and JRP wrote the manuscript.

ACKNOWLEDGEMENTS

This study was supported by funding from an NIDDK P20 grant (P20 DK103093) to J.D.B, R.B.W and J.R.P. We thank the Stanford Cancer Institute-supported core facilities providing services, including the Stanford Tissue Bank, the Stanford Functional Genomics Facility, and the Stanford Genome Sequencing Service Center.
REFERENCES


FIGURE LEGENDS

Figure 1. BPH transcriptional landscape and clinically-relevant stromal signature. (A) Heatmap of unsupervised clustering of normal prostate, BPH and BPH stromal nodules (samples color coded), across the 2,548 most variably-expressed genes. Transcript levels (median-centered log₂ RPKM values) are indicated by color-key, lower right. Select gene features (clusters) are annotated based on the expression of characteristic marker genes, including a stroma/myofibroblast feature (enframed on the heatmap) enriched in BPH over normal prostate and heterogeneous among BPH cases. LCM panel (left) indicates genes more highly expressed in laser microdissected BPH epithelium (orange) or BPH stroma (purple). Graph (right) illustrates correlation of gene expression features (moving average 51 genes) with BPH Symptom and Bother scores. Note, peak correlations reside within the core stromal feature. The dataset corresponding to the heatmap is available as Supplemental Table 2. (B) High expression of a 65-gene stromal signature is associated with elevated IPSS Symptom Score (left; strong trend) and Bother Score (right). Mean (red) and SD (blue) shown; P-values generated from 2-sided Student’s t-test. The 65 genes correspond to the core of the stromal gene feature (indicated in Figure 1A, lower right), with high expression defined as the top 50%. (C) A similarly derived 57-gene AR/Secretory signature showed no association with Symptom Score or Bother Score.

Figure 2. Genes differentially expressed in BPH vs. normal prostate. (A) Heatmap of genes with significant differential expression in BPH vs. normal prostate. Samples are clustered (see color key) and genes ordered by SAM score (t-statistic value). (B) Volcano plot (q-value vs. log₂ fold change) annotated with topmost genes differentially expressed in BPH. (C) Technical
validation of RNAseq results by qRT-PCR, confirming elevated expression of BMP5 and CXCL13 in BPH. Five BPH-normal pairs assayed once with technical quadruplicates. Graphed are mean (+/- 1SD) relative expression levels, normalized to GAPDH and compared against a reference sample (normal prostate sample 8791). Each data point represents one of up to 16 pairwise ratios (calculated from the quadruplicate test and reference values). Below, heatmap depiction of corresponding RNAseq (median-centered log2 RPKM) values.

Figure 3. BMP5 addition to cultured prostate cells supports role in BPH disease process.
(A) Addition of BMP5 (250ng/ml) to RWPE-1 prostate epithelial and WPMY-1 myofibroblast cells activates canonical SMAD pathway, demonstrated by 15-20 fold increased phospho-SMAD1/5/9 on western blot. (B) BMP5 addition (concentrations indicated) to RWPE-1 cells increases cell numbers. Mean (red) and SD (blue) shown. Multiplicity adjusted $P$-values generated from one-way ANOVA with post-hoc comparison to no BMP5 control (Dunnett test); *, $P<0.05$. Data are representative of three independent experiments, each done with two samples assayed per concentration. (C) Addition of BMP5 (250ng/ml) to RWPE-1 cells leads to dispersed cell growth (with fewer cell clusters) (left), and to increased transwell migration (right). Mean and SD shown; $P$-value generated from 2-sided Student’s t-test. Data are representative of two independent experiments, each done with three samples assayed per condition. (D, E) BMP5 addition to RWPE-1 and WPMY-1 cells generates a transcriptional response significantly enriched for BPH (vs. normal prostate) overexpressed genes (left), and with WPMY-1 cells for BPH stromal signature genes (center), but not AR/Secretory signature genes (right). GSEA enrichment score $P$-values are indicated.
Figure 4. IHC on prostate TMA identifies altered ESR1+ cell subset in BPH. (A-D)
Hematoxylin and Eosin (H&E) stains of representative normal prostate and BPH matched pair.
Original magnifications, 200X and 800X (insets). (E-H) CHGA immunostaining (marker of
neuroendocrine cells) shown for same case; note depletion of neuroendocrine cells in BPH. (I)
CHGA IHC scores across all TMA cases. Mean (red) and SD (blue) shown; P-values generated
from 2-sided Student’s t-test. (J-M) ESR1 immunostaining shown for same case; note depletion
of periglandular ESR1-positive fibroblast-like cells in BPH. (N) ESR1 IHC scores across all
TMA cases.

Figure 5. Single-cell RNAseq reveals altered cell subsets in BPH. (A) Two-dimensional
projection (t-SNE plot) of single cell transcriptomes stratifies prostate tissue cells (dots) into
distinct clusters, identifiable by characteristic expression of marker genes, including (B) KRT13
(basal epithelium) (red intensity scales to maximum log2 expression), (C) KLK3 (secretory
epithelium), and (D) DCN (fibroblasts). Additional cell type markers are shown in Supplemental
Figure 6. (E) Normal prostate tissue cell subset, illustrating expression of (F) CHGA
(neuroendocrine cells), (G) BMP5, and (H) CXCL13. (I) BPH tissue cell subset, illustrating
expression of (J) CHGA, (K) BMP5, and (L) CXCL13. Note relative depletion in BPH tissue of
CHGA-expressing neuroendocrine cells, and enrichment of BMP5/CXCL13 expressing
myofibroblasts. Insets magnify select cell clusters.

Figure 6. Whole exome sequencing reveals somatic SNVs in BPH nodules. (A) Integrated
Genome Viewer (IGV) display of mapped reads from WES of BPH and matched normal DNA
within the IL1B gene. Blue and red reads map to plus and minus strands; only a subset of
mapped reads (totals indicated) shown. Note, the IL1B SNV in BPH leads to a frameshift with early translational termination. (B) Similar IGV display depicts a nonsynonymous SNV in the TTN gene in a different BPH case. (C) Sanger sequencing validation of IL1B SNV. (D) Sanger sequencing validation of TTN SNV. See Supplemental Table 11 for additional SNVs validated by deep amplicon resequencing.
Figure 1. BPH transcriptional landscape and clinically-relevant stromal signature. (A) Heatmap of unsupervised clustering of normal prostate, BPH and BPH stromal nodules (samples color coded), across the 2,548 most variably-expressed genes. Transcript levels (median-centered log2 RPKM values) are indicated by color-key, lower right. Select gene features (clusters) are annotated based on the expression of characteristic marker genes, including a stroma/myofibroblast feature (enframed on the heatmap) enriched in BPH over normal prostate and heterogeneous among BPH cases. LCM panel (left) indicates genes more highly expressed in laser microdissected BPH epithelium (orange) or BPH stroma (purple). Graph (right) illustrates correlation of gene expression features (moving average 51 genes) with BPH Symptom and Bother scores. Note, peak correlations reside within the core stromal feature. The dataset corresponding to the heatmap is available as Supplemental Table 2. (B) High expression of a 65-gene stromal signature is associated with elevated IPSS Symptom Score (left; strong trend) and Bother Score (right). Mean (red) and SD (blue) shown; P-values generated from 2-sided Student’s t-test. The 65 genes correspond to the core of the stromal gene feature (indicated in Figure 1A, lower right), with high expression defined as the top 50%. (C) A similarly derived 57-gene AR/Secretory signature showed no association with Symptom Score or Bother Score.
Figure 2. Genes differentially expressed in BPH vs. normal prostate. (A) Heatmap of genes with significant differential expression in BPH vs. normal prostate. Samples are clustered (see color key) and genes ordered by SAM score (t-statistic value). (B) Volcano plot (q-value vs. log2 fold change) annotated with topmost genes differentially expressed in BPH. (C) Technical validation of RNAseq results by qRT-PCR, confirming elevated expression of BMP5 and CXCL13 in BPH. Five BPH-normal pairs assayed once with technical quadruplicates. Graphe are mean (+/- 1SD) relative expression levels, normalized to GAPDH and compared against a reference sample (normal prostate sample 8791). Each data point represents one of up to 16 pairwise ratios (calculated from the quadruplicate test and reference values). Below, heatmap depiction of corresponding RNAseq (median-centered log2 RPKM) values.
Figure 3. BMP5 addition to cultured prostate cells supports role in BPH disease process. (A) Addition of BMP5 (250 ng/ml) to RWPE-1 prostate epithelial and WPMY-1 myofibroblast cells activates canonical SMAD pathway, demonstrated by 15-20 fold increased phospho-SMAD1/5/9 on western blot. (B) BMP5 addition (concentrations indicated) to RWPE-1 cells increases cell numbers. Mean (red) and SD (blue) shown; P-values generated from 2-sided Student’s t-test. Data are representative of three independent experiments, each done with two samples assayed per concentration. (C) Addition of BMP5 (250 ng/ml) to RWPE-1 cells leads to dispersed cell growth (with fewer cell clusters) (left), and to increased transwell migration (right). Mean and SD shown. Multiplicity adjusted P-value generated from one-way ANOVA with post-hoc comparison to no BMP5 control (Dunnett test); *, P<0.05. Data are representative of two independent experiments, each done with three samples assayed per condition. (D, E) BMP5 addition to RWPE-1 and WPMY-1 cells generates a transcriptional response significantly enriched for BPH (vs. normal prostate) overexpressed genes (left), and with WPMY-1 cells for BPH stromal signature genes (center), but not AR/Secretory signature genes (right). GSEA enrichment score P-values are indicated.
Figure 4. IHC on prostate TMA identifies altered ESR1+ cell subset in BPH. (A-D) Hematoxylin and Eosin (H&E) stains of representative normal prostate and BPH matched pair. Original magnifications, 200X and 800X (insets). (E-H) CHGA immunostaining (marker of neuroendocrine cells) shown for same case; note depletion of neuroendocrine cells in BPH. (I) CHGA IHC scores across all TMA cases. Mean (red) and SD (blue) shown; P-values generated from 2-sided Student’s t-test. (J-M) ESR1 immunostaining shown for same case; note depletion of periglandular ESR1-positive fibroblast-like cells in BPH. (N) ESR1 IHC scores across all TMA cases.
Figure 5. Single-cell RNAseq reveals altered cell subsets in BPH. (A) Two-dimensional projection (t-SNE plot) of single cell transcriptomes stratifies prostate tissue cells (dots) into distinct clusters, identifiable by characteristic expression of marker genes, including (B) KRT13 (basal epithelium) (red intensity scales to maximum log2 expression), (C) KLK3 (secretory epithelium), and (D) DCN (fibroblasts). Additional cell type markers are shown in Supplemental Figure 6. (E) Normal prostate tissue cell subset, illustrating expression of (F) CHGA (neuroendocrine cells), (G) BMP5, and (H) CXCL13. (I) BPH tissue cell subset, illustrating expression of (J) CHGA, (K) BMP5, and (L) CXCL13. Note relative depletion in BPH tissue of CHGA-expressing neuroendocrine cells, and enrichment of BMP5/CXCL13 expressing myofibroblasts. Insets magnify select cell clusters.
Figure 6. Whole exome sequencing reveals somatic SNVs in BPH nodules. (A) Integrated Genome Viewer (IGV) display of mapped reads from WES of BPH and matched normal DNA within the IL1B gene. Blue and red reads map to plus and minus strands; only a subset of mapped reads (totals indicated) shown. Note, the IL1B SNV in BPH leads to a frameshift with early translational termination. (B) Similar IGV display depicts a nonsynonymous SNV in the TTN gene in a different BPH case. (C) Sanger sequencing validation of IL1B SNV. (D) Sanger sequencing validation of TTN SNV. See Supplemental Table 11 for additional SNVs validated by deep amplicon resequencing.