

Expressed sequence tag profiling identifies developmental and anatomic partitioning of gene expression in the mouse prostate

Denise E Abbott*, Colin Pritchard*, Nigel J Clegg*, Camari Ferguson*, Ruth Dumpit*, Robert A Sikes[†] and Peter S Nelson^{*‡}

Addresses: *Division of Human Biology, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, Seattle, WA 98109, USA.

[†]Laboratory for Cancer Ontogeny and Therapeutics, Department of Biological Sciences, University of Delaware, Newark, DE 19716, USA.

[‡]Division of Clinical Research, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, Seattle, WA 98109, USA.

Correspondence: Peter S Nelson. E-mail: pnelson@fhcrc.org

Published: 28 November 2003

Genome Biology 2003, 4:R79

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2003/4/12/R79>

Received: 11 September 2003

Revised: 28 October 2003

Accepted: 12 November 2003

© 2003 Abbott *et al.*; licensee BioMed Central Ltd. This is an Open Access article: verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article's original URL.

Abstract

Background: The prostate gland is an organ with highly specialized functional attributes that serves to enhance the fertility of mammalian species. Much of the information pertaining to normal and pathological conditions affecting the prostate has been obtained through extensive developmental, biochemical and genetic analyses of rodent species. Although important insights can be obtained through detailed anatomical and histological assessments of mouse and rat models, further mechanistic explanations are greatly aided through studies of gene and protein expression.

Results: In this article we characterize the repertoire of genes expressed in the normal developing mouse prostate through the analysis of 50,562 expressed sequence tags derived from 14 mouse prostate cDNA libraries. Sequence assemblies and annotations identified 15,009 unique transcriptional units of which more than 600 represent high quality assemblies without corresponding annotations in public gene expression databases. Quantitative analyses demonstrate distinct anatomical and developmental partitioning of prostate gene expression. This finding may assist in the interpretation of comparative studies between human and mouse and guide the development of new transgenic murine disease models. The identification of several novel genes is reported, including a new member of the β -defensin gene family with prostate-restricted expression.

Conclusions: These findings suggest a potential role for the prostate as a defensive barrier for entry of pathogens into the genitourinary tract and, further, serve to emphasize the utility of the continued evaluation of transcriptomes from a diverse repertoire of tissues and cell types.

Introduction

The normal function of the mammalian prostate gland is to enhance fertility by secreting buffers, proteins and protective agents that maintain sperm in a quiescent and intact state as they pass through the male and female reproductive tracts [1].

Much of the information pertaining to normal prostate physiology has been obtained through extensive developmental and biochemical analyses of the prostates of rodent species. Despite anatomical differences between rodents and humans, these studies have been instrumental for elucidating the

influence of androgens on prostate differentiation and growth, and for characterizing the protein and mineral constituents that comprise the unique prostate environment.

In humans, the prostate exhibits the distinctive attribute of sustained growth throughout life, a situation that contributes to both benign and malignant prostate pathologies [1]. Strikingly, the prevalence rates of benign prostate hypertrophy (BPH) and prostate carcinoma approach nearly 50% in American men by the age of 70 [2-4]. Despite extensive research efforts, the etiologies of these diseases remain poorly defined. In contrast to colon, skin and bladder epithelia, prostate epithelial cells are thought to be relatively better protected from environmental insults, and the cellular constituents exhibit low proliferation rates [5]. Androgenic hormones and hereditary factors influence both BPH and prostate carcinoma, but the specific mechanisms by which they alter cellular growth remain to be delineated.

As with other human health disorders, rodent models have been developed to aid in the scientific analysis of prostate diseases. Whilst early efforts focused extensively on the rat prostate [6,7] - in part due to the advantages of working with a relatively large gland - recent investigations have utilized the mouse increasingly, primarily as a result of the ease and power of manipulating the mouse genome [8]. The rodent prostate is comprised of four distinct lobes: ventral, anterior (also termed the coagulating gland), dorsal and lateral; the latter two are commonly grouped together and collectively referred to as the dorsolateral lobe [9]. These lobes are arranged circumferentially around the urethra and display characteristic patterns of ductal branching and secretory protein production. In contrast, the human prostate lacks a defined lobar architecture - it is organized in zones with distinct disease predispositions; carcinoma primarily develops in the peripheral zone and benign hypertrophy primarily occurs in the transition zone [10]. The anatomical and functional relationships between the rodent prostate lobes and the human prostate zones have not been definitively established, though it has been suggested that the human peripheral zone is most analogous to the rodent dorsolateral lobe based upon the observation that tumors induced in rodent prostates generally arise in these locations [7,11].

Although important insights pertaining to normal development and disease pathology can be obtained through detailed anatomical and histological assessments of mouse models, further mechanistic explanations are greatly aided through studies of gene and protein expression. In this context, the Cancer Genome Anatomy Project (CGAP) [12] and other large-scale sequencing efforts have sought to provide a comprehensive sequence and reagent set that encompasses genes expressed in diverse collections of human and mouse tissues [13]. However, a recent inventory of cDNA libraries and sequences archived in CGAP and the database of expressed sequence tags (ESTs) indicates that while 838 cDNA libraries

have been constructed from murine tissues and cell types, there is no mouse prostate representation (query 8.20.2003 in [12]). In this article we characterize the repertoire of genes expressed in the normal developing mouse prostate through the analysis of ESTs derived from mouse prostate cDNA libraries. The results of this analysis demonstrate distinct anatomical and developmental partitioning of gene expression, a finding that may assist in the interpretation of comparative studies between human and mouse, and further guide the evaluation and development of new transgenic murine disease models. The identification of several novel genes is reported, including a new member of the β -defensin gene family with prostate-restricted expression. This finding suggests a potential role for the prostate as a defensive barrier for entry of pathogens into the genitourinary tract.

Results

Mouse prostate transcriptome

To assess the diversity of gene expression in the mouse prostate, we constructed and characterized cDNA libraries representing distinct stages of prostate development and maturation that began with the urogenital sinus (E16.5), and continued with prostates from two day old (neonatal), ten day old, 20 day old, 35 day old (puberty), three month old (normal adult), and 14 month old (aged adult) animals (Figure 1). A total of 62,046 ESTs were generated, of which 50,562 passed stringent quality assessments. Of these, 4,299 ESTs annotated to the mitochondrial genome. Assembly of the remaining prostate ESTs into distinct clusters or transcription units (TUs) was performed using the sequence assembly program Phrap [14] and each TU was assigned an annotation by BLAST comparisons against sequences in public databases. In accordance with definitions used by the FANTOM and RIKEN teams in assembling a mouse transcriptome using full-length cDNAs [15], we use the term TU to designate a cluster of transcripts or ESTs that contain a common core of sequence information transcribed from a segment of the genome. A TU may be defined by a single cDNA sequence or by multiple overlapping ESTs that share a common core sequence. Assembly and annotation of the 46,263 mouse prostate ESTs produced 15,009 unique TUs. Of these, 4,550 TUs are composed of more than one EST and 10,459 TUs are represented by a single EST. Of the 4,550 TUs that contain more than one EST, 3,567 are represented by sequences derived from more than one mouse prostate library. All sequences are archived for public accession in the mouse prostate expression database (mPEDB) [16,17].

Gene expression alterations during mouse prostate development

To identify genes potentially involved with prostate development and maturation, we determined temporal expression changes by calculating transcript abundance levels in cDNA libraries constructed from different stages of prostate development. Pairwise comparisons of the datasets from each

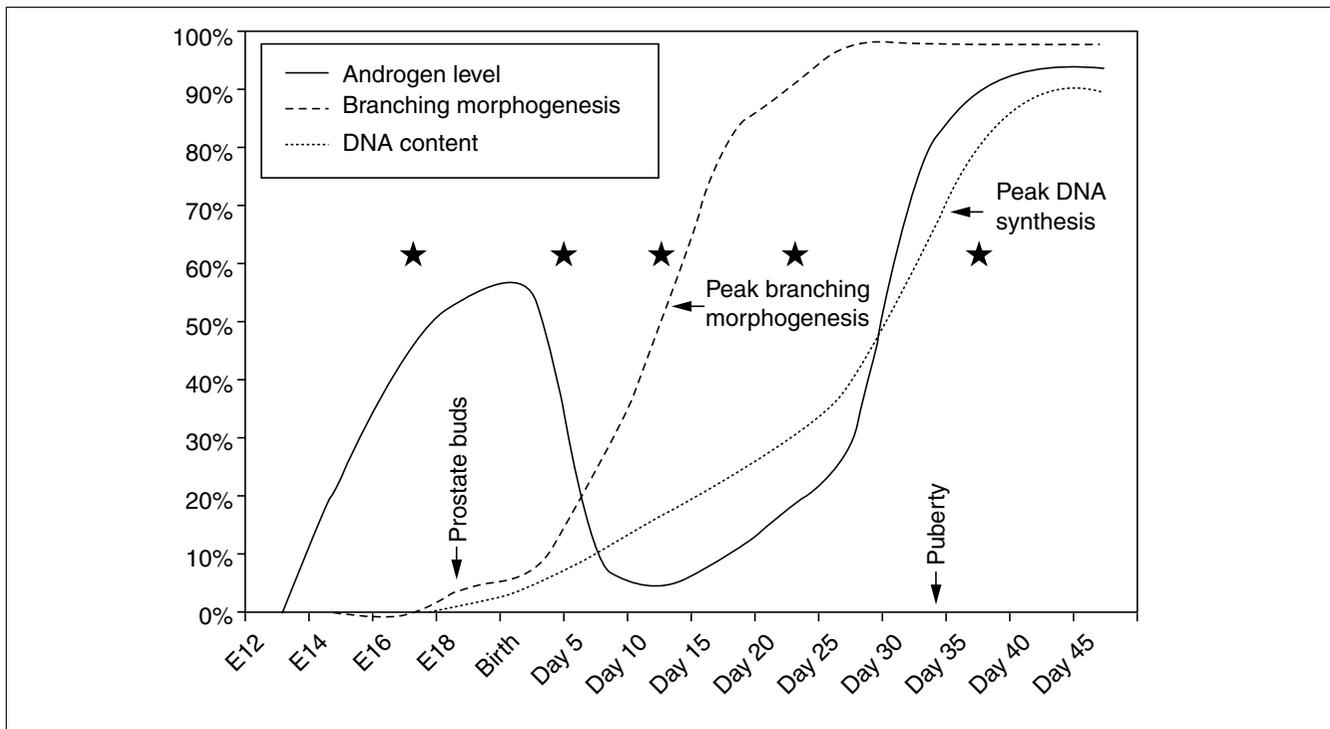


Figure 1

Temporal events in mouse prostate development. Androgen levels (solid line) rise beginning at day E12, fall shortly after birth, and rise again at puberty [65]. Branching morphogenesis (dashed line) begins at approximately day E17 with prostate budding, peaks at approximately day 10 (after birth), and is essentially complete before puberty by day 35 [23]. Peak DNA synthesis (dotted line), representing the development of the prostate epithelial and stromal cell mass, occurs at approximately day 35 [66]. cDNA libraries were constructed from prostate tissues obtained at defined points of prostate development (stars).

timepoint were performed using Audic and Claverie's analysis method [18]. At a significance level of $p = 0.001$, 285 genes were differentially expressed between one or more time points (see Additional data file 1, Table 1). Applying the Bonferroni correction for multiple comparisons identified a cohort of 69 genes with highly significant differential expression between one or more developmental stages. Hierarchical clustering of these genes demonstrated distinct temporal partitioning of gene expression with notable cohorts increasing over time and others decreasing over time. In Figure 2, two clusters of genes are highlighted which show progressive increases or decreases in expression over time from urogenital sinus to adulthood. Several of these genes encode proteins involved in specialized prostate secretory activity such as spermine binding protein (*Sbp*) [19] and serine protease inhibitor Kazal type 3 (*Spink3*) [20]. Several are also known to be regulated by androgenic hormones, such as probasin (*Pbsn*), and increase in expression following the onset of puberty [21]. Other TUs with temporal changes in expression have no currently identified functional roles in prostate development.

Gene expression compartmentalization between mouse prostate lobes

The rodent prostate gland is comprised of four distinct anatomical lobes: dorsal, lateral, ventral and anterior, each with unique ductal branching patterns and different responses to androgenic hormones (Figure 3) [22,23]. Studies of secretory products from rodent prostates have identified specific expression patterns that can be assigned to individual lobes [24-26]. To further explore the functional differences between mouse prostate lobes, we compared the profiles of genes expressed in each lobe and identified 34 genes that exhibited statistically significant differential expression. This group was reduced to 17 genes after correcting for multiple comparisons (Table 2). The expression patterns of these genes clearly support conclusions of prior anatomical and biochemical studies demonstrating functional heterogeneity between lobes of the gland. Genes highly expressed in the anterior prostate relative to the other lobes included onzin (alias placenta specific 8; *Plac8*), pancreatic family type A ribonuclease 1 (*RNase1*), and experimental autoimmune prostatitis antigen (*Eapa1*; a putative prostate transglutaminase), as well as several uncharacterized transcripts. Spermine binding protein (*Sbp*), and serine protease Kazal type 3 (*Spink3*) are relatively highly expressed in the ventral lobe, while transcripts with increased expression in the

Table 1**Characterized genes with enhanced prostate expression**

Symbol	UniGene description	Tissues	UniGene ID
<i>Syp2</i>	Seminal vesicle protein 2	Genitourinary; aorta and vein	Mm.1286
<i>Msmb</i>	β -microseminoprotein	Genitourinary	Mm.2540
<i>Galgt2</i>	UDP-N-acetyl-alpha-D-galactosamine	Colon; embryo	Mm.2807
<i>Nkx3-1</i>	NK-3 transcription factor locus 1 (<i>Drosophila</i>)	Trophoblast stem cell	Mm.3520
<i>Svs6</i>	Seminal vesicle secretion 6	Genitourinary	Mm.3787
<i>Sva</i>	Seminal vesicle antigen	Genitourinary	Mm.4119
<i>Pbsn</i>	Probasin	Genitourinary; bone	Mm.8034
<i>Defb2</i>	Defensin β 2	Head; epididymis	Mm.41981
<i>Sbp</i>	Spermine binding protein	Tongue; genitourinary; stomach	Mm.46428
<i>B3galt1</i>	B 1 3-galactosyltransferase polypeptide 1	Medulla oblongata; ganglion	Mm.57041
<i>ErbB3</i>	Erythroblastic leukemia viral oncogene homolog 3	Tumor; inner ear	Mm.57112
<i>Hoxd12</i>	Homeo box D12	Embryo; forelimb	Mm.57124
-	11 kDa secreted protein precursor	Whole skin	Mm.71887
<i>Mlt7</i>	Mixed lineage-leukemia translocation to 7 homolog	Tumor; genitourinary; spleen	Mm.88827
<i>Smk2</i>	Sperm motility kinase 2	Lung	Mm.88851
<i>Svs7</i>	Seminal vesicle protein secretion 7	Genitourinary	Mm.99349
<i>Fhl3</i>	Four and a half LIM domains 3	Placenta	Mm.100241
<i>Pappa</i>	Pregnancy-associated plasma protein A	Parthenogenote; tumor; genitourinary	Mm.103481
<i>Svs3</i>	Seminal vesicle secretion 3	Genitourinary; adipose	Mm.118769
-	EST similar to Acrosomal vesicle protein 1	Genitourinary; thymus; placenta	Mm.118804
<i>Svs2</i>	Seminal vesicle protein secretion 2	Genitourinary; aorta and vein	Mm.143501
<i>Amd2</i>	S-adenosylmethionine decarboxylase 2	Blastocyst	Mm.195848
<i>Acac</i>	Acetyl-coenzyme A carboxylase	Whole brain; branchial arches	Mm.196688
<i>Muc10</i>	Mucin 10 submandibular gland salivary mucin	Salivary gland	Mm.200411
<i>Lamc2</i>	Laminin gamma2 chain	Inner ear	Mm.213197
<i>Birc1f</i>	Baculoviral IAP repeat-containing 1f	Lymph node; colon	Mm.218759
<i>Tnep1</i>	Thrombospondin N-terminal domain	Whole brain	Mm.221237

dorsolateral prostate include receptor activity modifying protein 2 (*Ramp2*), a transcript with similarity to acrosomal vesicle protein-1, and several uncharacterized transcripts. We determined the lobe-specific expression patterns of several transcripts using quantitative PCR (QPCR), and found that the EST and QPCR methods produced qualitatively similar results (Figure 4).

Wubah *et al.* have reported the cloning and tissue distribution of a transcript termed ventral prostate predominant 1 (*Vpp1*) that was identified through mRNA differential display between different mouse prostate lobes [27]. QPCR and virtual expression analysis confirmed that the expression of *Vpp1* is primarily localized in the ventral prostate, with lower but detectable levels of expression in the dorsolateral and anterior prostate lobes (Figure 4).

Figure 2 (see following page)

Differential gene expression during mouse prostate development. Statistical analysis of transcript abundance levels in stages of mouse prostate development identified 69 differentially expressed genes ($p = 0.001$). Shown are cohorts with progressing or regressing levels over time. The intensity scale represents the fold-change in expression normalized relative to the lowest abundance measurement.

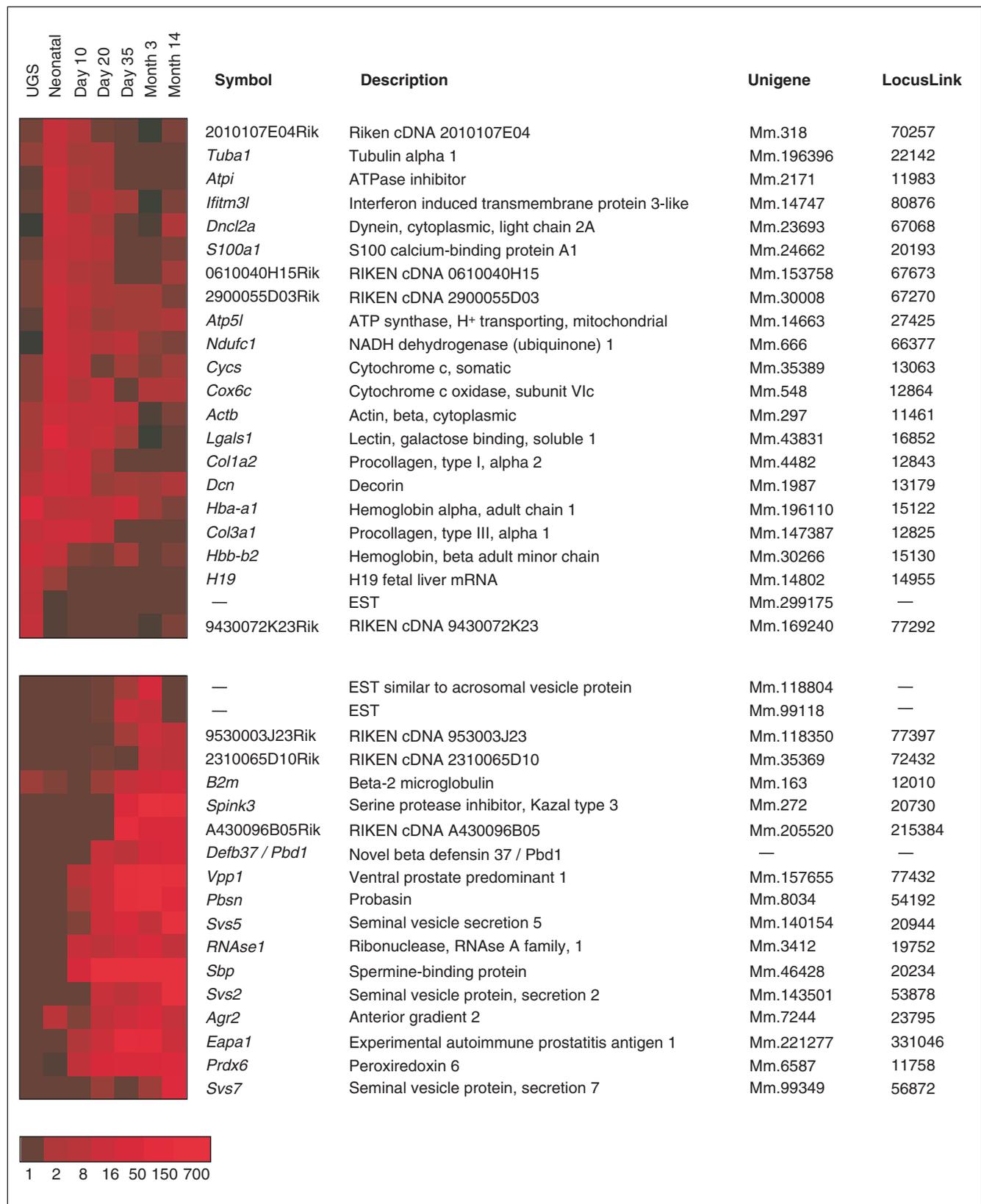


Figure 2 (see legend on previous page)

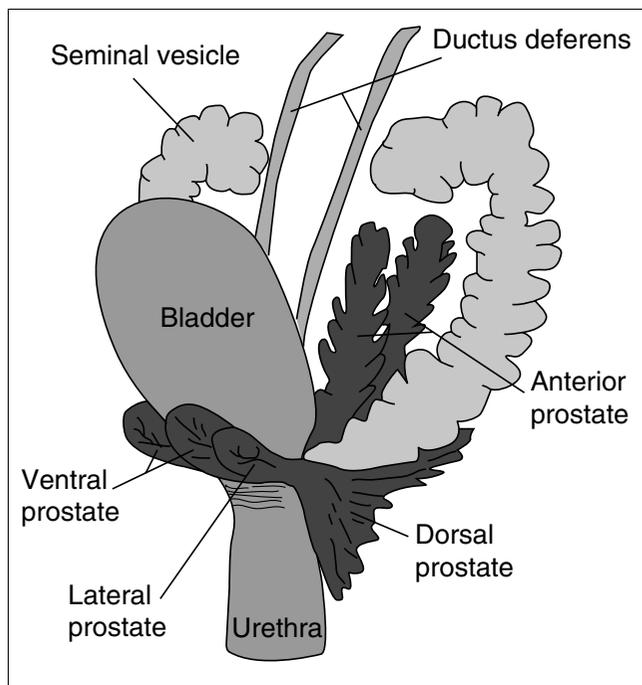


Figure 3
A schematic diagram of the adult mouse genitourinary tract (lateral view), with the anatomically-distinct prostate lobes highlighted in gray. Reproduced with permission from [67,68].

Our EST results demonstrated that the probasin transcript was most highly expressed in the anterior prostate with lower - but still high - transcript abundance levels in the dorsolateral lobes, and minimal expression in the ventral lobe (Figure 4). QPCR quantitated the highest probasin expression in the dorsolateral lobes followed by the anterior prostate, with lowest levels in the ventral lobe. These results differ from studies of expression localization using the rat probasin promoter to drive transgene expression in mouse prostate epithelium where the highest levels were observed in the lateral and dorsal lobes, with lower expression in the ventral prostate and very low to absent expression in the anterior prostate [11,28]. Subsequent immunohistochemical and Western analyses with antibodies recognizing the mouse probasin protein demonstrated high levels of expression in the anterior as well as the dorsolateral mouse prostate lobes [29]. These results suggest that the rat probasin promoter may confer slightly different cellular specificity compared with the native mouse promoter, or that transgenic constructs alter the normal regional distribution of probasin expression.

Genes with expression enhanced or restricted to the mouse prostate

The comparative analysis of cDNA libraries representing the repertoire of genes expressed in the mouse prostate provided an opportunity to identify genes whose expression is enhanced or restricted to the mouse prostate relative to other normal tissues. The human prostate gland expresses several

transcripts and corresponding proteins in a highly tissue-restricted manner including prostate specific antigen (PSA) [30], human glandular kallikrein 2 (hK2) [31], prostase/ KLK4 [32] and prostate specific membrane antigen (PSMA) [33]. Studies of these proteins have provided insights into normal prostatic function, mechanisms of hormone-regulated gene expression and the development of prostate pathology. To identify transcripts preferentially expressed in the mouse prostate, we assigned each prostate TU to a UniGene cluster and then determined the tissue sources of all ESTs comprising the cluster. In total, 776 prostate TUs were mapped to UniGenes containing ESTs derived from three or fewer other tissues (see Additional data file 2). Only 28 of these TUs represent characterized genes (Table 1). Included among the 28 are NKX3.1 and probasin, both of which are known to have prostate-restricted patterns of expression [29,34]. Most of the prostate-enhanced TUs are represented by uncharacterized ESTs and full-length cDNAs sequenced by the RIKEN mouse transcript sequencing project [15]. Eighty TUs contained ESTs from at least two different prostate libraries. Of these, eight represent spliced gene products based on sequence comparisons with the mouse draft genome sequence. An additional 20 TUs had interesting features, such as high sequence conservation with human sequences, and may represent orthologous genes.

Identification of a gene encoding a putative prostate-specific β -defensin

One TU, originally designated TU23348 after the original library clone identifier, was found to be expressed in multiple mouse prostate cDNAs libraries, lack representation in the UniGene or dbEST sequence databases, and exhibit sequence similarities to genes in the β -defensin family. We reasoned that since TUs specifically originating from the mouse prostate were not represented in the public sequence databases, TU23348 could represent a β -defensin with expression restricted to the prostate. To evaluate this possibility, we measured the relative abundance of TU23348 transcripts in multiple mouse tissues by QPCR. These results confirmed high levels of TU23348 expression in the prostate, with very low levels in testis and skeletal muscle, and no detectable transcripts in a wide range of other tissues (Figure 5a). A multiple tissue mRNA blot confirmed the RT-PCR findings (Figure 5b), and northern analysis identified a transcript size of approximately 0.4 kb. Northern analysis and QPCR also indicated that TU23348 expression is further compartmentalized within the mouse prostate having highest expression levels in the dorsolateral and ventral lobes, with very low to absent expression in the anterior prostate (Figure 5c).

A full-length TU23348 transcript was obtained through assemblies of mouse prostate cDNAs and by 5' and 3' rapid amplification of cDNA ends (RACE) reactions using mouse prostate cDNA as a template. The coding sequence comprises 219 nucleotides encoding a predicted polypeptide of 73 amino acids (Figure 6a). The nucleotide sequence comprises a 72 nt

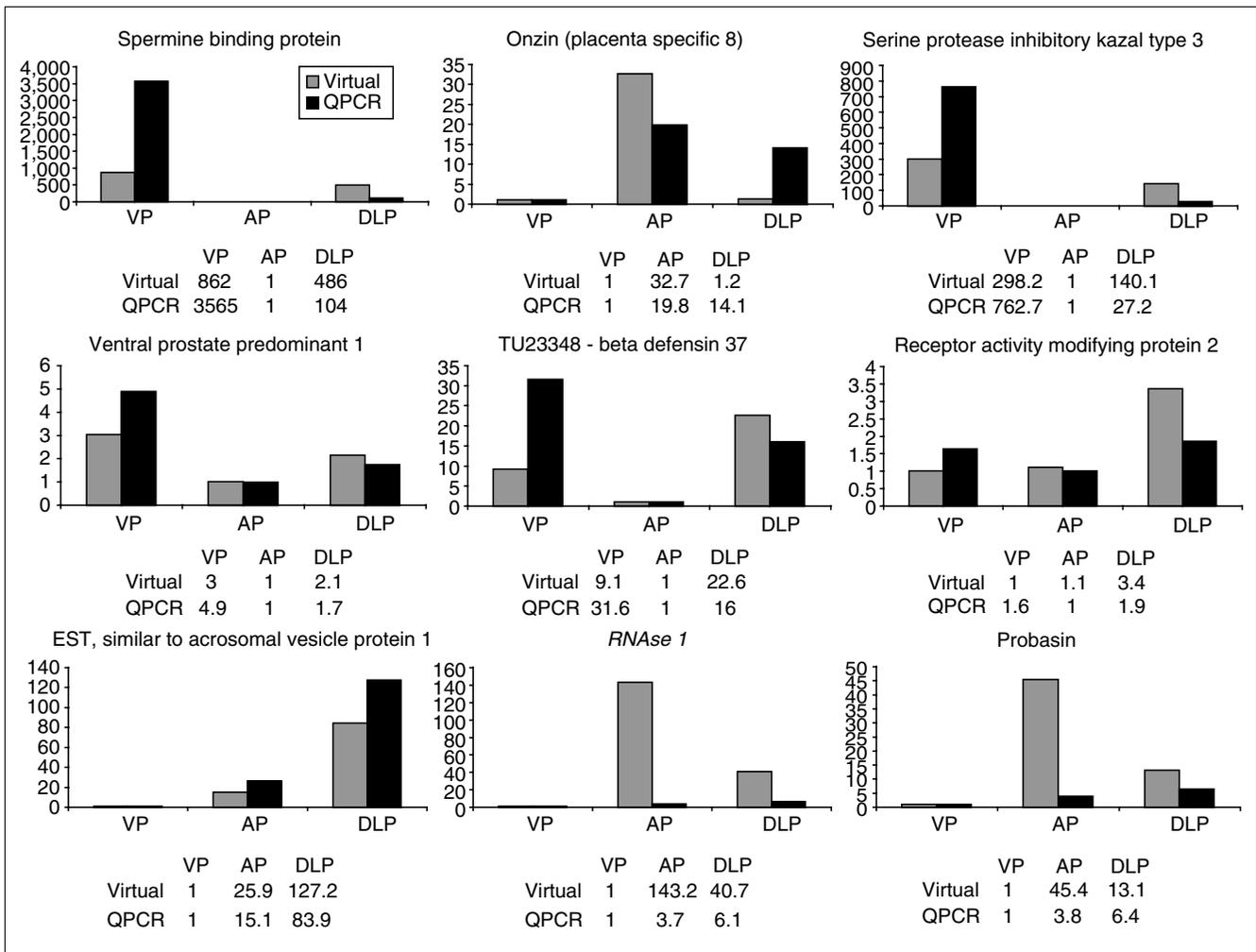


Figure 4 Comparative analysis of fold change of gene expression in adult prostate lobes using quantitative PCR (QPCR) (dark grey bars) and virtual EST measurements (light grey bars).

5' UTR and 111 nt 3' UTR with putative polyadenylation sequences located 99, 72 and 51 nt upstream of the polyA tail. Protein sequence alignments with members of the mouse and rat β -defensin family demonstrate a high level of sequence similarity including conservation of the majority of the cysteine and positively-charged residues that are predicted to modulate antimicrobial activity (Figure 6) [35,36]. Alignments of the TU23348 transcript with the draft mouse genome sequence indicates that TU23348 is located on chromosome 8 and is comprised of two exons; the first encoding 19 amino acids and the second 54 amino acids. The exons are separated by a 7.3 kb intron (Figure 6). The gene is located between two other members of the β -defensin family, β -defensin 1 and 2. We named this gene *Pbd1* for prostate β -defensin 1. Subsequently, in accordance with the Human Genome Organization (HUGO) nomenclature committee, this gene carries the official designation of *Defb37* for

defensin beta 37 as it is the 37th β -defensin reported. The sequence is represented in GenBank under accession number AY387658.

Discussion

The physiological and biochemical features of a particular tissue or cell type represent the complex endpoint of interactions between specific cohorts of expressed genes and the environment. The identification of the complete set of genes expressed in the mouse has been the focal point of intensive efforts involving the sequence analysis of hundreds of cDNA libraries derived from tissues at different stages of mouse development [15]. The ultimate success of this approach depends on the specific organs, tissues and cell types selected for sampling. Many transcripts will be expressed in a tissue- or cell type-restricted manner, while others are expressed

Table 2**Genes with relative differential expression between mouse prostate lobes ($p = 0.001$ using the Bonferroni correction) with normalized EST numbers**

Symbol	Description	UniGene	Species ID	ESTs per lobe		
				VP	AP	DLP
<i>Sbp</i>	Spermine binding protein	Mm.46428	6313	3618 ¹	0 ¹	2076 ¹
<i>Spink3</i>	Serine protease inhibitor Kazal type 3	Mm.272	6296	1242 ¹	0 ¹	594 ¹
<i>Vpp1</i>	Ventral prostate predominant 1	Mm.157655	5394	583 ¹	200 ^{1,2}	420 ²
<i>Ramp2</i>	Receptor (calcitonin) activity modifying protein 2	Mm.218611	4672	42 ¹	48	157 ¹
<i>Defβ37</i>	Novel prostate β-defensin 37	-	10164	38	0 ¹	97 ¹
A430096B05Rik	RIKEN cDNA A430096B05	Mm.205520	4722	33 ¹	147 ¹	78
<i>Pbsn</i>	Probasin	Mm.8034	5239	23 ¹	1100 ¹	311 ¹
<i>Eapa1</i>	Experimental autoimmune prostatitis antigen 1	Mm.221277	5446	18 ¹	379 ¹	168 ¹
<i>Prdx6</i>	Peroxiredoxin 6	Mm.6587	2712	15 ^{1,2}	187 ¹	86 ²
<i>B2m</i>	β-2 microglobulin	Mm.163	5263	3 ^{1,2}	113 ¹	57 ²
<i>Svs2</i>	Seminal vesicle protein secretion 2	Mm.143501	1342	2 ¹	187 ^{1,2}	2 ²
9530003J23Rik	RIKEN cDNA 9530003J23	Mm.118350	5050	2 ¹	170 ^{1,2}	0 ²
<i>RNAse1</i>	Ribonuclease 1 pancreatic	Mm.3412	4606	0 ¹	248 ¹	69 ¹
<i>Plac8</i>	Onzin (placenta-specific 8)	Mm.34609	5767	0 ¹	57 ¹	2
<i>Svs5</i>	Seminal vesicle secretion 5	Mm.140154	506	0 ¹	31	44 ¹
-	ESTs similar to acrosomal vesicle protein 1	Mm.118804	2570	0 ¹	26 ²	143 ^{1,2}

^{1,2} Statistically significant gene expression comparisons between lobes are denoted by shared superscripts. VP, ventral prostate; AP, anterior prostate; DLP, dorsolateral prostate.

only in response to specific stimuli or at precise stages of development [15,37]. Indeed, despite the comprehensive analysis and assembly of more than two million mouse cDNAs and ESTs into about 37,000 distinct TUs, a subsequent report focusing on genes expressed in murine macrophages and dendritic cells identified more than 300 high quality TUs that had not been identified in the comprehensive study [37].

Despite the prominence of diseases affecting the prostate gland and the development of numerous mouse models to evaluate prostate pathology, there are no reports of comprehensive gene expression studies of the normal mouse prostate. In addition, normal mouse prostate EST resources are not included in public sequence data repositories. In part this may reflect the difficulties encountered in working with organs of very small size, as well as the presence of high levels of endogenous RNAses in the mouse prostate gland. The analysis of the prostate transcriptome reported here is based on the assembly of >50,000 ESTs derived from 14 cDNA libraries constructed from distinct anatomical regions and developmental stages. A total of 15,009 TUs were defined by 4,550 clusters and 10,459 singleton sequences. These TUs annotated to 9,882 known genes leaving 5,127 as potentially novel uncharacterized transcripts. Of the uncharacterized TUs, 683 exhibited an open reading frame of at least 100 amino acids. The remaining 4,444 TUs may comprise 5' or 3' untranslated

regions of protein-coding genes, or possibly non-coding RNAs. This number represents ~30% of all distinct prostate TUs, a number in good agreement with conclusions reached from analyses of the RIKEN mouse EST dataset in which 11,665 TUs or about 35% of all TUs were determined to be non-coding transcripts [15]. While the functional significance of these mRNAs has not been determined, non-coding RNAs may be of importance in normal and pathological processes affecting the prostate. Two human genes, *PCGEM1* and *DD3*, appear to produce non-coding transcripts that are overexpressed in human prostate carcinoma relative to benign tissue counterparts [38,39].

In the mouse, structures comprising the urogenital sinus develop beginning at day 13 post conception with prostate buds forming late on day 17, a developmental stage roughly correlating with the maximum production of testosterone by the fetal testis [1]. At birth, androgen receptor activity is detected primarily in mesenchymal cell types comprising the prostate stroma with a subsequent transition to expression in the secretory epithelium following their differentiation from basal epithelium. To identify genes potentially involved with prostate development, we determined temporal expression changes by measuring transcript abundance levels in cDNA libraries constructed from distinct stages of maturation. Genes involved in specialized prostate secretory function demonstrated temporal expression increases coinciding with

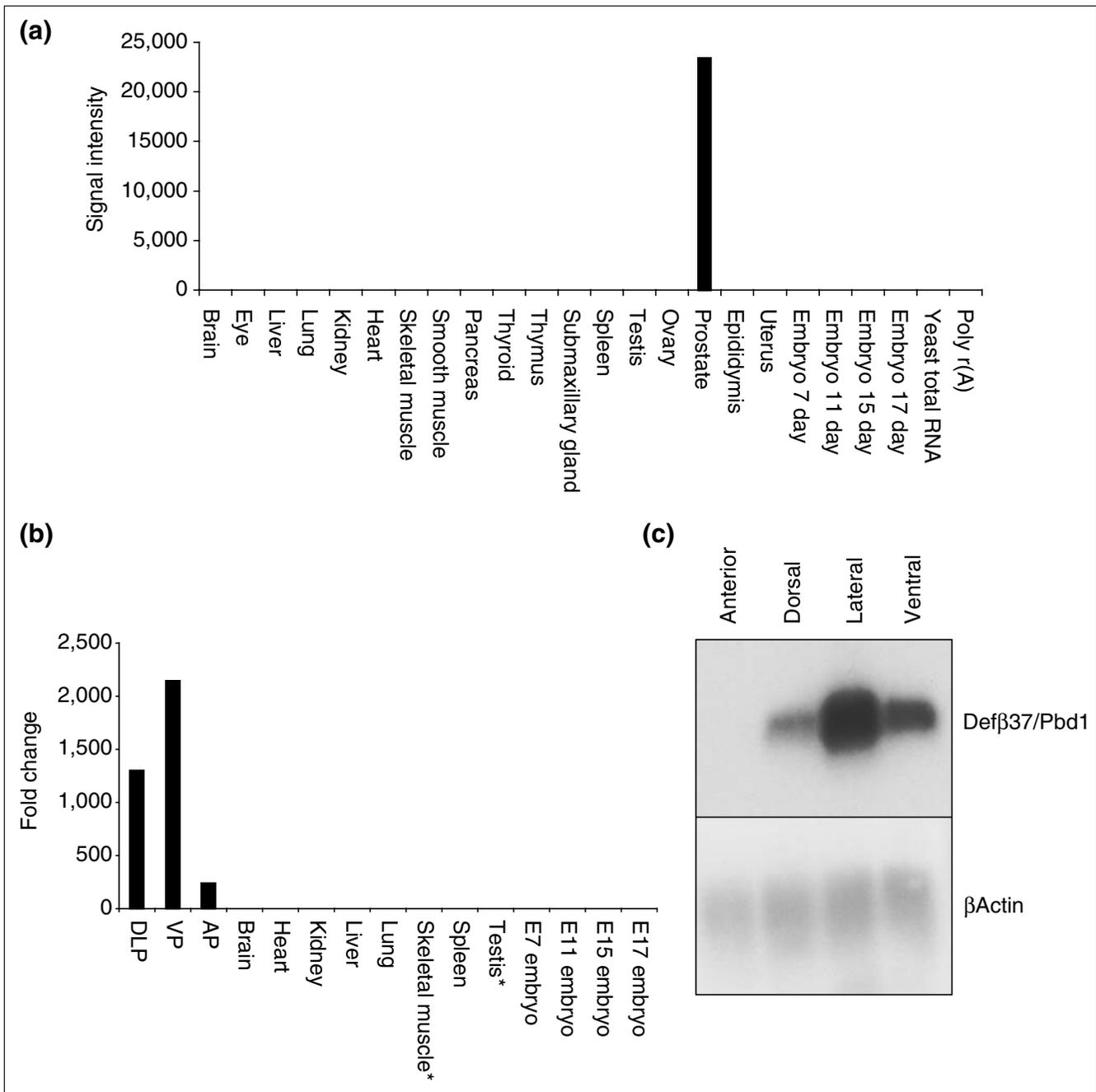


Figure 5 Prostate-specific expression of a novel β -defensin gene (*Defbeta37/Pbd1*). **(a)** Dot blots comprised of mRNAs derived from multiple normal mouse tissues were hybridized with probes encoding the *Defbeta37/Pbd1* gene, exposed to a phosphorimage screen, and quantitated. Tissue signal intensities are reported in arbitrary units above background intensity. **(b)** Prostate-specific expression of a novel β -defensin gene (*Defbeta37/Pbd1*). QPCR analysis of *Defbeta37/Pbd1* expression in multiple normal mouse tissues and in dissected mouse prostate lobes. Expression levels reflect the relative fold-differences in transcript abundance between tissue samples. Very low levels of *Defbeta37/Pbd1* expression not visible on this plot were detectable in skeletal muscle and testis (asterisks). **(c)** Prostate-specific expression of a novel β -defensin gene (*Defbeta37/Pbd1*). Northern analysis of *Defbeta37/Pbd1* expression in dissected individual mouse prostate lobes.

prostate maturation and the production of testosterone at puberty. These genes include several that encode seminal fluid proteins such as *Svs2*, *Svs5* and *Sbp*. The expression of probasin, a gene known to be regulated by androgens,

followed a similar pattern of expression. Transcripts encoding the antioxidant protein peroxiredoxin 6 (*Prdx6*) (alias antioxidant protein 2: *Aop2*) also increased with age. *Prdx6* is involved in the redox regulation of the cell and can reduce

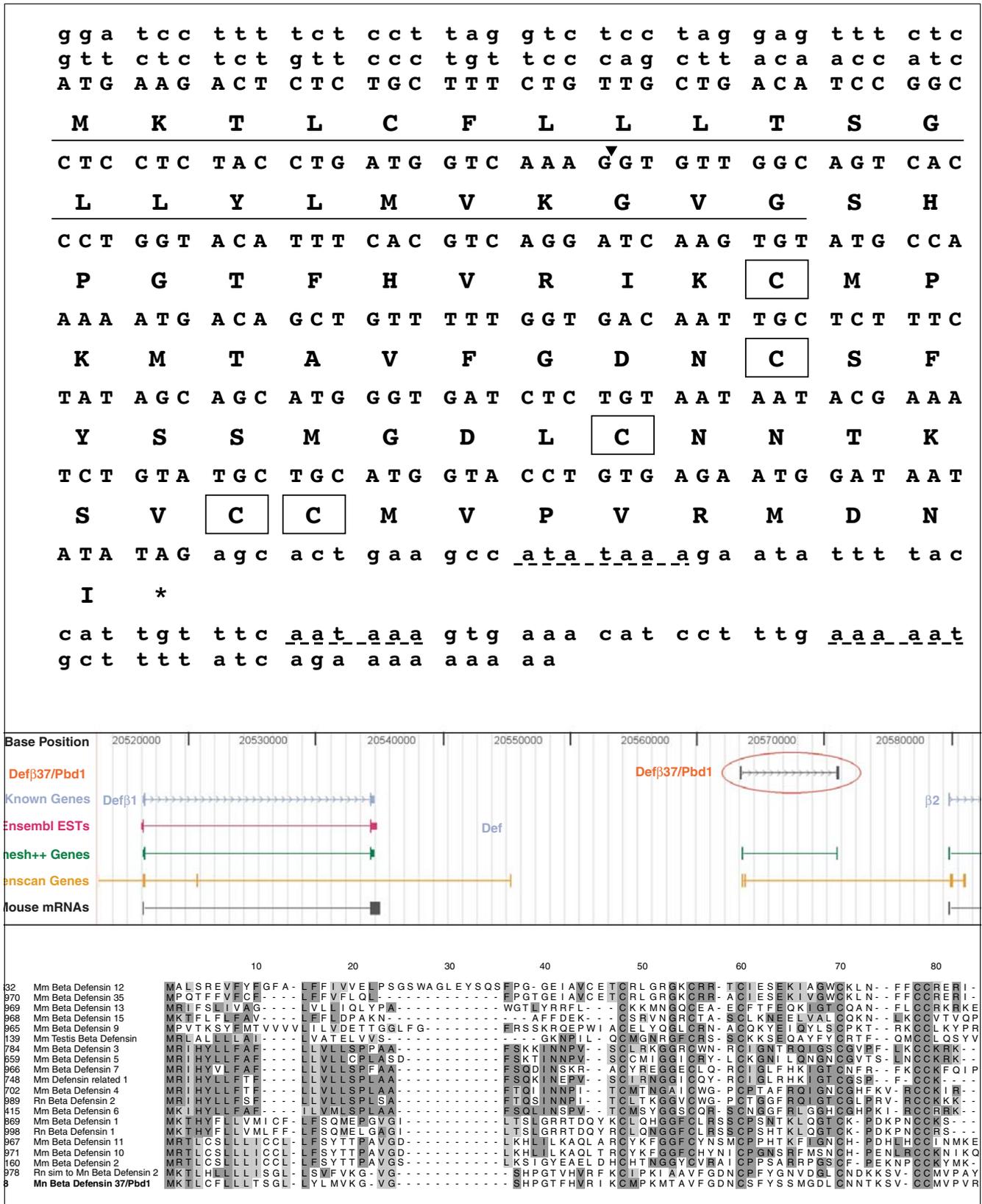


Figure 6 (see legend on next page)

Figure 6 (see previous page)

Sequence analysis of *Defβ37/Pbd1*. **(a)** The full-length mouse *Defβ37/Pbd1* sequence was obtained through the assembly of mouse prostate ESTs and RACE reactions. The open reading frame (capital small font) comprises 219 nt encoding a putative protein of 73 amino acids. Three potential polyadenylation signals are located 99, 72 and 51 nt upstream of the polyA sequence (dashed underline). The *Defβ37/Pbd1* gene is predicted to have two exons (splice site denoted by inverted triangle). A leader sequence of 22 amino acids (underline) is predicted based upon comparison with other members of the β-defensin family. Cysteine residues conserved with other β-defensin genes are boxed. **(b)** Alignment of *Defβ37/Pbd1* with the draft mouse genome sequence using BLAT. *Defβ37/Pbd1* aligns with sequence on mouse chromosome 8 and is spliced into two exons separated by a 7.3 kb intron. Genscan and Fgenesh predict a gene corresponding to *Pbd1*. No public mouse mRNAs or ESTs align to the *Defβ37/Pbd1* sequence. *Defβ37/Pbd1* is flanked by β-defensins 1 and 2. **(c)** Multiple protein sequence alignment with *Defβ37/Pbd1* and β-defensin family members. Residues conserved in a majority of sequences are boxed in dark gray. Residues strongly similar to conserved sequences are boxed in light gray. Cysteine residues predicted to be involved in β-defensin activity are conserved between family members.

H₂O₂, short chain fatty acids and phospholipid hydroperoxides [40]. The expression of such redox-modulating enzymes may assist in modulating the unusual high zinc/high citrate biochemical phenotype described in the human prostate microenvironment [41]. α-tubulin, dynein and other genes involved in motility and the cytoskeleton were maximally expressed in neonatal prostate with diminished expression correlating with gland maturation. Genes involved in energy metabolism such as *Cox6c* and *ATP5l* were similarly highly expressed early in prostate development. Surprisingly, transcripts encoding α- and β-globin were measured at high levels early in development, suggesting that silencing mechanisms restricting globin expression to red blood cell lineages are yet to be operative [42]. Other studies have identified globin gene expression in non-erythroid cell types in response to specific stimuli [43], and it has been hypothesized that globin proteins may function not only to carry oxygen, but may also participate in other cellular processes [43,44]. Interestingly, one of the first mouse models of prostate carcinoma was developed inadvertently using the human γ-globin promoter to drive expression of the SV40T antigen [45,46]. The experimental objective of the γ-globulin/SV40T transgenic mouse was to produce a model of erythroleukemia, but tumors of the prostate, adrenal cortex and brown adipose resulted from the genetic manipulation, indicating that the human γ-globin promoter exerts activity in prostate epithelial cells.

We further characterized one novel prostate TU, now designated *Defβ37/Pbd1*, which exhibits high sequence homology with the β-defensin gene family. Alignment of the *Defβ37/Pbd1* sequence with the draft mouse genome sequence placed it between two members of the β-defensin family, β-defensin 1 and 2, located on chromosome 8. No mouse ESTs annotated to this region, suggesting that *Defβ37/Pbd1* was restricted in expression to the prostate. We confirmed the prostate-localized expression of *Defβ37/Pbd1* by QPCR and northern analysis. The products of the defensin gene family function as cationic antimicrobial peptides and serve as components of the phagocytic and epithelial innate host defense system [47]. A recent report indicates that the β-defensin family now comprises at least 43 members in the mouse, organized into five clusters located on chromosomes 1, 2 (two clusters), 8 and 14, with orthologs present on corresponding chromosomal regions in the human genome [35]. While most of the

characterized β-defensins are expressed in many tissues, several have been identified with expression restricted to organs such as the epididymis and testis [48]. Other peptides with antimicrobial features and sequence similarity to the defensins have also been identified, such as the rat Bin1b protein that is expressed exclusively in the epididymis [49]. In addition to antimicrobial activity, the defensins have also been shown to influence non-specific cytotoxicity, membrane permeability and chemotaxis. They also function as sperm-immobilizing agents and have cytotoxicity toward oocytes and preimplantation embryos [47,50]. As with other regions of the genitourinary tract, the prostate gland is subject to infection by bacteria and other pathogens resulting in the clinical diagnosis of prostatitis. Thus, *Defβ37/Pbd1* may have a role in preventing infectious disease within the prostate, or if secreted into seminal fluid it may function to modulate fertility.

This study represents the first global analysis of transcript expression in the mouse prostate, and provides both a virtual as well as a physical resource for further studies of prostate development, physiology and pathology. Comparative studies of mouse and human prostate gene expression may assist in determining the functional equivalents between the lobar and zonal anatomies of the mouse and human prostates, respectively. Exploiting the regulatory mechanisms dictating prostate-specific and androgen-regulated gene expression may also assist in the development of new transgenic models and provide for a more complete understanding of hormonal regulation. The identification of previously uncharacterized transcripts such as *Defβ37/Pbd1* serves to emphasize the need to evaluate transcriptomes from a diverse repertoire of tissues and cell types in order to catalog completely the genes expressed in mammalian species.

Methods

cDNA library construction and DNA sequence analysis

Male C57Bl6J mice (Jackson Labs, Maine, USA) of defined ages were sacrificed in accordance with institutional protocols. Whole urogenital sinus (UGS) or individual prostate lobes were dissected, pooled and snap frozen in liquid nitrogen. Tissues were homogenized using a Polytron PT-MR-2100 rotor stator homogenizer (Kinematica, Switzerland) and total RNA was purified using Qiagen RNeasy purification

system (Qiagen, Valencia, CA, USA). For libraries made from whole prostate, equal amounts of RNA from each lobe were pooled prior to cDNA synthesis. Approximately 1 µg of total RNA from each tissue source was used in first strand cDNA synthesis reactions using the SMART cDNA library construction kit in accordance with the manufacturer's protocol (Clontech, CA, USA). Second strand cDNA was synthesized and amplified through 18-21 cycles of the PCR, digested with SfiI and directionally ligated into the lambdaTriplEx2 phagemid. Phagemids were converted to pTriplEx2 plasmids in the BM25.8 *Escherichia coli* strain. A total of 14 cDNA libraries were constructed. From each library, between 3,000 and 15,000 bacterial clones were picked into 384-well plates using a Q-pix robot (Genetix, OR, USA) and grown overnight at 37°C in LB media supplemented with 8% glycerol. We used 384-pin-replicators (Genetix) to transfer directly to 10 µl PCR reactions containing Triplex2 forward PCR primer (5' CTCGGGAAGCGCGCCATTGTGTTGGT 3') and Triplex2 reverse PCR primer (5' ATACGACTCACTATAGGGCGAATTGGCC 3'). PCR reactions were incubated with 1.5 units of exonuclease I (Epicentre) and 0.5 units of shrimp alkaline phosphatase (USB) for 15 minutes at 37°C to remove primers and nucleotide triphosphates. Pin-replicators were used to transfer ~0.2 µl of purified PCR product to 5 µl fluorescence-based sequencing reactions containing 0.6 µM Triplex2 5'Seq primer (5' CTCGGGAAGCGCGCCATTGTGTTGGT 3'), 0.5 µL Big Dye Terminator (ABI), 4 mM Tris pH 9.0 and 1 mM EDTA. One urogenital sinus library (UGSo2) was derived from approximately 500 Balb/c strain fetal mice and was synthesized using both oligo-dT and random hexamer priming followed by size-selection for cDNAs greater than 500 basepairs. These cDNAs were ligated to BstXI and EcoRI adaptors, cloned into pcDNAII plasmids (Invitrogen, CA, USA) and sequenced as described above but with the following primers: VN26 PCR (5' TTTCCAGTCACGACGTTGTA 3'), VN27 PCR (5' GTGAGCGGATAACAATTTTCAC 3'), M13R 5'Seq (5' GGAAACAGCTATGACCATG 3'). Detailed library descriptions are available at the mouse prostate expression database website [16,17].

Assembly and annotation of ESTs derived from the mouse prostate

Each cDNA was sequenced once to produce an EST averaging 500 bp in length. Each EST was screened for sequence comprising vectors, *E. coli*, repetitive elements and low complexity regions using the RepeatMasker algorithm. Sequence quality was determined using Phred basecalling software [51]. Following quality assessment, ESTs comprised of >100 unmasked nucleotides with 80/100 bases demonstrating a quality score exceeding 20 were included in the analysis. Each EST was compared against the mouse reference sequence database using the Basic Local Alignment Search Tool (BLAST) [52] and assigned a corresponding gene annotation if the BLAST score exceeded 1,100. A BLAST score of 1,100 represents a minimum match of 555 nucleotides with 100% identity, or a longer sequence match with relatively few

mismatches or gaps in the alignment; a stringency suitable for matching ESTs against curated gene sequences. The remaining ESTs not matching a reference sequence with high stringency were assembled into gene clusters using the Phrap assembly software with minimum match length parameter set to 60 and minimum score parameter set to 90 [53]. Consensus sequences were determined and compared with sequences in the UniGene, GenBank, and dbEST databases [54] using BLAST. ESTs that received a BLAST score of 200 or better were assigned a database annotation. The remaining ESTs were classified as unannotated. The number of ESTs from each library corresponding to the same annotation were enumerated and used for subsequent comparative determinations of gene expression measurements.

ESTs and EST assemblies not receiving annotations in UniGene, GenBank or dbEST were further evaluated for characteristics supporting their classification as genes as opposed to artifacts such as genomic DNA. Sequences represented by only one EST were not further evaluated. ESTs represented in ≥ 2 different libraries were compared against the mouse genome assembly (February 2003 freeze) using the BLAT algorithm to assess for splicing and gene predictions [55,56]. These sequences were further evaluated for protein coding regions by conceptual translations of six possible reading frames using a custom perl script. Sequences with greater than 200 amino acids of translated sequence in any reading frame were analyzed on the BLOCKS server [57], run with default settings, to identify possible motifs. Conceptual translations were also checked for protein similarity using the PSI- and PHI-BLAST algorithms [58].

Virtual analysis of differential gene expression

To identify genes with restricted or enhanced expression in the prostate relative to other mouse tissues, we searched the UniGene database with each prostate TU and selected those UniGene clusters containing sequences derived from two or fewer non-prostate tissues. For the purposes of this screen the terms 'urinary bladder', 'vesicular gland', 'Wolffian duct includes surrounding region', 'testicles', 'Wolffian duct and mullerian duct' were grouped into a 'genitourinary' category. In addition, 'tumor biopsy sample', 'tumor gross tissue', 'pooled lung tumors', 'tumor metastatic to mammary', 'infiltrating ductal carcinoma', 'spontaneous tumor metastatic to mammary', and 'stem cell origin' were grouped under 'tumor'.

To identify genes differentially expressed in prostate development or between mouse prostate lobes, we utilized the program Identification of Differentially Expressed Genes 6 test statistics (IDEG6) [59,60]. IDEG6 performs a normalization calculation to adjust for libraries with different numbers of ESTs. Six different algorithms are applied to the datasets to determine those TUs with statistical differences in abundance levels between two or more libraries. Romualdi *et al.* identified the Audic-Claverie approach [18] as the best statistical method for identifying differentially expressed genes in

pairwise comparisons [60], and we report the results of this analysis using a significance level of $p = 0.001$ with the Bonferroni correction to adjust for multiple comparisons. The DLP01, CG02, VP01 and VP02 libraries were evaluated in a pairwise fashion for the lobular comparison. Libraries included in the developmental pairwise comparison were UGS01, UGS02, NEONATAL, DAY10, DAY20, DAY35, a pooled library for MONTH3 (represented by combined TUs from the lobular comparison) and MONTH14. Genes and ESTs exhibiting differential expression between distinct developmental time points were grouped by hierarchical clustering using the Cluster software [61]. Clusters were visualized using Treeview [61].

Unannotated, putatively novel sequences determined to be differentially expressed between prostate lobes were further evaluated using PSI-BLAST (Position-Specific Iterated BLAST) and PHI-BLAST (Pattern-Hit Initiated BLAST) [58]. Nucleotide sequences were translated using the ExpASY Translate program [62] to identify open reading frames. ESTs were also compared against the assembled mouse genome (February 2003 freeze) through the University of Santa Cruz's BLAST-Like Alignment Tool (BLAT) [56] to identify spliced sequences and predicted genes.

Quantitative polymerase chain reaction

Quantitative PCR analysis of selected genes was performed to confirm the virtual expression results. Approximately 2 ng of cDNA from each prostate lobe was used as a template for PCR reactions using buffers and instructions in accordance with the manufacturer's protocol (Applied Biosystems Inc., CA, USA). Reactions contained 1X SYBR Green master mix (Applied Biosystems, Inc.), and 0.3 μ M of oligonucleotide primer pairs designed to amplify the following genes: probasin (*Pbsn*): Forward- 5' GGTCATCATCTCCTGCTCA 3', Reverse- 5' AGGCCGTCATCTTCTTTTT 3' (79 bp amplicon); spermine binding protein (*Sbp*): Forward- 5' CCCA-CATGCAGAGCCAGAAA 3', Reverse- 5' ATCCGCATGCCCTTGAGTTG 3' (95 bp amplicon); serine protease inhibitor, Kazal type 3 (*Spink3*): Forward- 5' TAT-AGTTCTTCTGGCTTTTGC 3', Reverse- 5' TCTATGCGTTTC-CTGTTTTCA 3' (246 bp amplicon); onzin (alias placenta-specific 8 (*Plac8*)): Forward- 5' TTCTGTCCTGTTTGCTCT-GTG 3', Reverse- 5' TCATGGCTCTCCTCTGTTA 3' (61 bp amplicon); receptor (calcitonin) activity modifying protein 2 (*Ramp2*): Forward- 5' CTCATCCTTCCCACAGACCT 3', Reverse- 5' TGTGTCGTGAGTCCCCTTTG 3' (61 bp amplicon); ventral prostate predominant 1 (*Vpp1*): Forward- 5' TGCTGTCTGTCTGCTTCTG 3', Reverse- 5' CCATACTTATT-GTTTCTCCTTTC 3' (126 bp amplicon); ribonuclease, RNase A Family, 1 (*RNase1*): Forward- 5' AAGTCCCTCATTCTGTT-TCCA 3', Reverse- 5' TATCCCGCGTTTCATCATTT 3' (166 bp amplicon); EST similar to acrosomal vesicle protein 1: Forward- 5' TGTTCCCTAGGCTCTCACTGC 3', Reverse- 5' CCAAGAGTAGCAACAAGAGG 3' (52 bp amplicon); prostate β -defensin 1 (*Pbd1*): Forward- 5' GATCAAGTGTAT-

GCCAAAATG 3', Reverse- 5' TTTATATGGCTTCAGT-GCTCTA 3' (149 bp amplicon). Amplifications were performed using a protocol of 95°C for 30 sec, 60°C for 30 sec, and 72°C for 30 sec for 40 cycles on the ABI 7700 sequence detector while recording fluorescence measurements. Each reaction was performed in triplicate and cycle numbers were normalized to a parallel ribosomal protein S16 control as previously described [63]. To determine the tissue distribution of *Pbd1*, quantitative PCR was performed using cDNAs derived from multiple mouse tissues according to the supplier's protocol (Invitrogen).

Northern and dot blot analysis

Ten μ g of total RNA from each prostate lobe and other mouse tissues were fractionated on 1.2% agarose denaturing gels and transferred to nylon membranes by a capillary method [64]. The mouse multiple tissue and master blots were obtained from Clontech, CA, USA. Blots were hybridized with *Def β 37/Pdb1* cDNA probes labeled with [α -³²P]dCTP by random priming using the Rediprime II random primer labeling system (Amersham, CA, USA) according to the manufacturer's protocol. Blots were stripped and re-probed for β -actin as a loading control. Filters were imaged and quantitated by using a phosphor-capture screen and ImageQuant software (Amersham, CA, USA).

Additional data files

The additional data files are available with this article and at [17]: A table showing the genes that are differentially expressed between developmental timepoints at a significance level of $p = 0.001$, without the correction for multiple comparisons (Additional data file 1), and a table of genes represented by more than two prostate ESTs and expressed in fewer than two tissues in addition to the prostate (Additional data file 2).

Acknowledgements

We thank Cory Abate-Shen for critical reading of the manuscript and helpful suggestions. This work was supported by grant DK63919 (to R.A.S.) and grants R01DK59125, R01DK65204, CA97186 and a Damon Runyon Scholar Award (to P.S.N.).

References

- Marengo SR: **Prostate Physiology and Regulation**. In: *Advanced Therapy of Prostate Disease* Edited by: Resnick MI, Thompson IM. Hamilton/London: BC Decker, Inc; 2000:92-117.
- Garraway WM, Collins GN, Lee RJ: **High prevalence of benign prostatic hypertrophy in the community**. *Lancet* 1991, **338**:469-471.
- Yatani R, Kusano I, Shiraishi T, Hayashi T, Stemmermann GN: **Latent prostatic carcinoma: Pathological and epidemiological aspects**. *Jpn J Clin Oncol* 1989, **19**:319-326.
- Haas GP, Sakr WA: **Epidemiology of prostate cancer**. *CA Cancer J Clin* 1997, **47**:273-287.
- Berges RR, Vukanovic J, Epstein JI, CarMichel M, Cisek L, Johnson DE, Veltri RW, Walsh PC, Isaacs JT: **Implication of cell kinetic changes during the progression of human prostatic cancer**. *Clin Cancer Res* 1995, **1**:473-480.
- Navone NM, Logothetis CJ, von Eschenbach AC, Troncoso P: **Model**

- systems of prostate cancer: uses and limitations.** *Cancer Metastasis Rev* 1998, **17**:361-371.
7. Pollard M, Luckert PH: **Autochthonous prostate adenocarcinomas in Lobund-Wistar rats: a model system.** *Prostate* 1987, **11**:219-227.
 8. Abate-Shen C, Shen MM: **Mouse models of prostate carcinogenesis.** *Trends Genet* 2002, **18**:S1-S5.
 9. Jesik CJ, Holland JM, Lee C: **An anatomic and histologic study of the rat prostate.** *Prostate* 1982, **3**:81-97.
 10. McNeal JE: **The zonal anatomy of the prostate.** *Prostate* 1981, **2**:35-49.
 11. Greenberg NM, DeMayo F, Finegold MJ, Medina D, Tilley WD, Aspinall JO, Cunha GR, Donjacour AA, Matusik RJ, Rosen JM: **Prostate cancer in a transgenic mouse.** *Proc Natl Acad Sci USA* 1995, **92**:3439-3443.
 12. **The Cancer Genome Anatomy Project - cDNA library finder** [<http://cgap.nci.nih.gov/Tissues/LibraryFinder>]
 13. Schaefer C, Grouse L, Buetow K, Strausberg RL: **A new cancer genome anatomy project web resource for the community.** *Cancer J* 2001, **7**:52-60.
 14. **The Phred/Phrap/Consed system home page** [<http://www.phrap.org>]
 15. Okazaki Y, Furuno M, Kasukawa T, Adachi J, Bono H, Kondo S, Nikaido I, Osato N, Saito R, Suzuki H et al.: **Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs.** *Nature* 2002, **420**:563-573.
 16. Nelson PS, Pritchard C, Abbott D, Clegg N: **The human (PEDB) and mouse (mPEDB) Prostate Expression Databases.** *Nucleic Acids Res* 2002, **30**:218-220.
 17. **Mouse Prostate Expression Database** [<http://www.mpedb.org>]
 18. Audic S, Claverie JM: **The significance of digital gene expression profiles.** *Genome Res* 1997, **7**:986-995.
 19. Mills JS, Needham M, Parker MG: **Androgen regulated expression of a spermine binding protein gene in mouse ventral prostate.** *Nucleic Acids Res* 1987, **15**:7709-7724.
 20. Mills JS, Needham M, Parker MG: **A secretory protease inhibitor requires androgens for its expression in male sex accessory tissues but is expressed constitutively in pancreas.** *EMBO J* 1987, **6**:3711-3717.
 21. Spence AM, Sheppard PC, Davie JR, Matuo Y, Nishi N, McKeehan WL, Dodd JG, Matusik RJ: **Regulation of a bifunctional mRNA results in synthesis of secreted and nuclear probasin.** *Proc Natl Acad Sci USA* 1989, **86**:7843-7847.
 22. Timms BG, Mohs TJ, Didio LJ: **Ductal budding and branching patterns in the developing prostate.** *J Urol* 1994, **151**:1427-1432.
 23. Sugimura Y, Cunha GR, Donjacour AA: **Morphogenesis of ductal networks in the mouse prostate.** *Biol Reprod* 1986, **34**:961-971.
 24. Hayashi N, Sugimura Y, Kawamura J, Donjacour AA, Cunha GR: **Morphological and functional heterogeneity in the rat prostatic gland.** *Biol Reprod* 1991, **45**:308-321.
 25. Imasato Y, Onita T, Moussa M, Sakai H, Chan FL, Koropatnick J, Chin JL, Xuan JW: **Rodent PSP94 gene expression is more specific to the dorsolateral prostate and less sensitive to androgen ablation than probasin.** *Endocrinology* 2001, **142**:2138-2146.
 26. Kinbara H, Cunha GR: **Ductal heterogeneity in rat dorsal-lateral prostate.** *Prostate* 1996, **28**:58-64.
 27. Wubah JA, Fischer CM, Rolfzen LN, Khalili M, Kang J, Green JE, Bieberich CJ: **Ventral prostate predominant I, a novel mouse gene expressed exclusively in the prostate.** *Prostate* 2002, **51**:21-29.
 28. Greenberg NM, DeMayo FJ, Sheppard PC, Barrios R, Lebovitz R, Finegold M, Angelopoulos R, Dodd JG, Duckworth ML, Rosen JM et al.: **The rat probasin gene promoter directs hormonally and developmentally regulated expression of a heterologous gene specifically to the prostate in transgenic mice.** *Mol Endocrinol* 1994, **8**:230-239.
 29. Johnson MA, Hernandez I, Wei Y, Greenberg N: **Isolation and characterization of mouse probasin: an androgen-regulated protein specifically expressed in the differentiated prostate.** *Prostate* 2000, **43**:255-262.
 30. Rittenhouse HG, Finlay JA, Mikolajczyk SD, Partin AW: **Human Kallikrein 2 (hK2) and prostate-specific antigen (PSA): two closely related, but distinct, kallikreins in the prostate.** *Crit Rev Clin Lab Sci* 1998, **35**:275-368.
 31. Grauer LS, Charlesworth MC, Saedi MS, Finlay JA, Liu RS, Kuus-Richel K, Young CY, Tindall DJ: **Identification of human glandular kallikrein hK2 from LNCaP cells.** *J Androl* 1996, **17**:353-359.
 32. Nelson PS, Gan L, Ferguson C, Moss P, Gelinis R, Hood L, Wang K: **Molecular cloning and characterization of prostase, an androgen-regulated serine protease with prostate-restricted expression.** *Proc Natl Acad Sci USA* 1999, **96**:3114-3119.
 33. Gregorakis AK, Holmes EH, Murphy GP: **Prostate-specific membrane antigen: current and future utility.** *Semin Urol Oncol* 1998, **16**:2-12.
 34. Bieberich CJ, Fujita K, He WW, Jay G: **Prostate-specific and androgen-dependent expression of a novel homeobox gene.** *J Biol Chem* 1996, **271**:31779-31782.
 35. Schutte BC, Mitros JP, Bartlett JA, Walters JD, Jia HP, Welsh MJ, Casavant TL, McCray PB Jr: **Discovery of five conserved beta-defensin gene clusters using a computational search strategy.** *Proc Natl Acad Sci USA* 2002, **99**:2129-2133.
 36. Wu Z, Hoover DM, Yang D, Boulegue C, Santamaria F, Oppenheim JJ, Lubkowski J, Lu W: **Engineering disulfide bridges to dissect antimicrobial and chemotactic activities of human beta-defensin 3.** *Proc Natl Acad Sci USA* 2003, **100**:8880-8885.
 37. Wells CA, Ravasi T, Sultana R, Yagi K, Carninci P, Bono H, Faulkner G, Okazaki Y, Quackenbush J, Hume DA et al.: **Continued discovery of transcriptional units expressed in cells of the mouse mononuclear phagocyte lineage.** *Genome Res* 2003, **13**:1360-1365.
 38. Srikanth V, Zou Z, Petrovics G, Xu L, Augustus M, Davis L, Livezey JR, Connell T, Sesterhenn IA, Yoshino K et al.: **PCGEM1, a prostate-specific gene, is overexpressed in prostate cancer.** *Proc Natl Acad Sci USA* 2000, **97**:12216-12221.
 39. Bussemakers MJ, van Bokhoven A, Verhaegh GW, Smit FP, Karthaus HF, Schalken JA, Debruyne FM, Ru N, Isaacs WB: **DD3: a new prostate-specific gene, highly overexpressed in prostate cancer.** *Cancer Res* 1999, **59**:5975-5979.
 40. Manevich Y, Sweitzer T, Pak JH, Feinstein SI, Muzykantov V, Fisher AB: **I-Cys peroxiredoxin overexpression protects cells against phospholipid peroxidation-mediated membrane damage.** *Proc Natl Acad Sci USA* 2002, **99**:11599-11604.
 41. Hochachka PW, Rupert JL, Goldenberg L, Gleave M, Kozlowski P: **Going malignant: the hypoxia-cancer connection in the prostate.** *Bioessays* 2002, **24**:749-757.
 42. Hardison R: **Hemoglobins from bacteria to man: evolution of different patterns of gene expression.** *J Exp Biol* 1998, **201**:1099-1117.
 43. Liu L, Zeng M, Stamler JS: **Hemoglobin induction in mouse macrophages.** *Proc Natl Acad Sci USA* 1999, **96**:6643-6647.
 44. Giardina B, Messana I, Scatena R, Castagnola M: **The multiple functions of hemoglobin.** *Crit Rev Biochem Mol Biol* 1995, **30**:165-196.
 45. Perez-Stable C, Altman NH, Brown J, Harbison M, Cray C, Roos BA: **Prostate, adrenocortical, and brown adipose tumors in fetal globin/T antigen transgenic mice.** *Lab Invest* 1996, **74**:363-373.
 46. Perez-Stable C, Altman NH, Mehta PP, Defetos LJ, Roos BA: **Prostate cancer progression, metastasis, and gene expression in transgenic mice.** *Cancer Res* 1997, **57**:900-906.
 47. Yang D, Biragyn A, Kwak LW, Oppenheim JJ: **Mammalian defensins in immunity: more than just microbicidal.** *Trends Immunol* 2002, **23**:291-296.
 48. Yamaguchi Y, Nagase T, Makita R, Fukuhara S, Tomita T, Tominaga T, Kurihara H, Ouchi Y: **Identification of multiple novel epididymis-specific beta-defensin isoforms in humans and mice.** *J Immunol* 2002, **169**:2516-2523.
 49. Li P, Chan HC, He B, So SC, Chung YW, Shang Q, Zhang YD, Zhang YL: **An antimicrobial peptide gene found in the male reproductive system of rats.** *Science* 2001, **291**:1783-1785.
 50. Sawicki W, Mystkowska ET: **Contraceptive potential of peptide antibiotics.** *Lancet* 1999, **353**:464-465.
 51. Ewing B, Hillier L, Wendl MC, Green P: **Base-calling of automated sequencer traces using phred. I. Accuracy assessment.** *Genome Res* 1998, **8**:175-185.
 52. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215**:403-410.
 53. Gordon D, Abajian C, Green P: **Consed: a graphical tool for sequence finishing.** *Genome Res* 1998, **8**:195-202.
 54. **NCBI Databases** [<http://www.ncbi.nlm.nih.gov/Database/index.html>]
 55. Kent WJ: **BLAT - the BLAST-like alignment tool.** *Genome Res* 2002, **12**:656-664.
 56. **UCSC Genome Bioinformatics** [<http://genome.ucsc.edu/>]
 57. Pietrokovski S, Henikoff JG, Henikoff S: **The Blocks database - a system for protein classification.** *Nucleic Acids Res* 1996, **24**:197-200.
 58. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lip-

- man DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic Acids Res* 1997, **25**:3389-3402.
59. **Comparative evaluation of statistical tests for the detection of differentially expressed genes in multiple tag sampling experiments** [<http://telethon.bio.unipd.it/bioinfo/IDEG6>]
60. Romualdi C, Bortoluzzi S, D'Alessi F, Danieli GA: **IDEG6: a web tool for detection of differentially expressed genes in multiple tag sampling experiments.** *Physiol Genomics* 2003, **12**:159-162.
61. Eisen MB, Spellman PT, Brown PO, Botstein D: **Cluster analysis and display of genome-wide expression patterns.** *Proc Natl Acad Sci USA* 1998, **95**:14863-14868.
62. **ExpASY Proteomics tools** [<http://us.expasy.org/tools>]
63. Pritchard CC, Hsu L, Delrow J, Nelson PS: **Project normal: defining normal variance in mouse gene expression.** *Proc Natl Acad Sci USA* 2001, **98**:13266-13271.
64. Sambrook J, Fritsch EF, Maniatis T: *Molecular Cloning: A Laboratory Manual* Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press; 1989.
65. Corpechot C, Baulieu EE, Robel P: **Testosterone, dihydrotestosterone and androstenediols in plasma, testes and prostates of rats during development.** *Acta Endocrinol (Copenh)* 1981, **96**:127-135.
66. Sugimura Y, Cunha GR, Donjacour AA, Bigsby RM, Brody JR: **Whole-mount autoradiography study of DNA synthetic activity during postnatal development and androgen-induced regeneration in the mouse prostate.** *Biol Reprod* 1986, **34**:985-995.
67. Bhatia-Gaur R, Donjacour AA, Sciavolino PJ, Kim A, Desai N, Young P, Norton CR, Gridley T, Cardiff RD, Cunha GR et al.: **Roles for Nkx3.1 in prostate development and cancer.** *Genes Dev* 1999, **13**:966-977.
68. Cunha GR, Donjacour AA, Cooke PS, Mee S, Bigsby RM, Higgins SJ, Sugimura Y: **The endocrinology and developmental biology of the prostate.** *Endocr Rev* 1987, **8**:338-362.