The Prostate 60:227–239 (2004)

Characterization and Comparative Analyses of Transcriptomes From the Normal and Neoplastic Human Prostate

Nigel Clegg,[1] Denise Abbott,[1] Camari Ferguson,[1] Roger Coleman,[1] and Peter S. Nelson[1,2]*

[1]Division of Human Biology, Fred Hutchinson Cancer Research Center, Seattle, Washington
[2]Division of Clinical Research, Fred Hutchinson Cancer Research Center, Seattle, Washington

BACKGROUND. The prostate gland is a highly specialized organ with functional attributes that serve to enhance the fertility of mammalian species. Pathological processes affecting the prostate include benign prostate hypertrophy and prostate carcinoma; diseases that account for major morbidity and mortality in middle-aged and elderly men. To facilitate studies of biological processes uniquely represented in the prostate and assess molecular alterations associated with prostate carcinoma, we sought to establish the diversity of gene expression in the normal and neoplastic prostate through the compilation and analysis of a prostate transcriptome.

METHODS. We assembled and annotated ESTs derived from prostate cDNA libraries that were either produced in our laboratory or available from public sequence repositories such as CGAP, dbEST, and Unigene. Determinations of differential gene expression between the normal prostate, other normal tissues, and neoplastic prostate tissues was performed using statistical algorithms. Confirmation of differential expression was performed by quantitative PCR and Northern analysis.

RESULTS. A total of 99,448 high-quality ESTs were assembled and annotated to produce a prostate transcriptome comprised of 24,580 distinct TUs. Comparative analyses of gene expression levels identified 61 TUs with exclusive expression in the prostate and 45 TUs with high levels of expression in the prostate relative to at least 25 other normal tissues ($P > 0.99$). Comparative analyses of ESTs derived from neoplastic prostate tissues identified 75 genes with dysregulated expression in cancer ($P > 0.99$).

CONCLUSIONS. The human prostate expresses a diverse repertoire of genes that reflect a functionally complex organ. The identification of genes with prostate-restricted or enhanced expression may provide additional insights into the biochemical processes that interact to form the developmental, signaling, and metabolic pathways of the normal and neoplastic gland. *Prostate 60: 227–239, 2004.* © 2004 Wiley-Liss, Inc.

KEY WORDS: **prostate; cDNA; EST; database; PEDB; cancer; transcriptome**

Abbreviations: EST, expressed sequence tag; TU, transcriptional unit; cDNA, copy DNA; SAGE, serial analysis of gene expression; QPCR, quantitative polymerase chain reaction; TGMA4, transglutaminase; MSMB, β-microseminoprotein; SMG1, seminogelin 1; SMG2, seminogelin 2; ARSDR1, retinol dehydrogenase 11; KLK2, kallikrein 2; KLK3, kallikrein 3; ACPP, prostate acid phosphatase; NKX3A, NKX3 transcript homolog A; TMPRSS2, transmembrane protease serine 2; TRPM7, transient receptor potential cation channel M7; ARHU, ras homolog gene family member U; RP19, ribosomal protein L19; EEF2, eukaryotic translation elongation factor 2; FABP5, fatty acid binding protein 5; LOC54543, 6.2 kd protein; CKB, brain creatine kinase; FBLN5, fibulin 5; UQCRC2, ubiquinol-cytochrome c reductase core protein 2; QP-C, 9.5 kDa ubiquinone binding protein; NBL1, neuroblastoma suppression of tumorogenicity 1; ACTG2, actin gamma 2; SAT, spermidine/spermine N1-acetyltransferase; ALCAM, activated leucocyte cell adhesion molecule; TRG, T cell receptor gamma locus; TRGLN, transgelin; LAMB2, laminin beta 2; RBBP7, retinoblastoma binding protein 7; PP1R7, protein phosphatase 1 regulatory subunit number 7; PCANAP5, prostate cancer associated protein 5.

Grant sponsor: DOD; Grant number: PC991274; Grant sponsor: NIH; Grant numbers: CA97186, CA85859, DK65204; Grant sponsor: Damon Runyon Cancer Research Foundation (to P.S.N., a scholar award).

*Correspondence to: Peter S. Nelson, MD, Division of Human Biology, Mailstop D4-100, 1100 Fairview Avenue North, Seattle, WA. E-mail: pnelson@fhcrc.org

Received 24 September 2003; Accepted 25 November 2003
DOI 10.1002/pros.20055
Published online 12 February 2004 in Wiley InterScience (www.interscience.wiley.com).

INTRODUCTION

Determining the complete sequence of the human genome and identifying each encoded gene have been major goals of recent biological research. As the biochemical and physiological features of a particular cell type represent the complex interactions between expressed genes, large-scale efforts have also focused on characterizing the subset of genes that distinguish particular tissues and pathological processes. The drive to establish complete qualitative and quantitative expression profiles led to the development of high-throughput approaches for sequencing transcript tags and the establishment of sequence archives such as the database of expressed sequence tags (dbEST) [1], the cancer genome anatomy project (CGAP) [2], and the serial analysis of gene expression (SAGE) repository [3].

Collections of ESTs have been used in qualitative studies to establish transcriptomes reflecting the diversity of gene utilization in a particular organism, tissue, or cell type (e.g., http//cgap.nci.hih.gov). To facilitate such comprehensive descriptions, the term TU has been used to designate a unique segment of the genome from which transcripts are generated [4]. A given TU is defined by clusters of transcripts that contain a common core of genetic information [4]. In the context of EST libraries, computational algorithms are used to group the tags based on sequence homology or common database annotations such that each EST is incorporated into a specific TU. Assemblies of this sort are represented by databases such as Unigene [5], the TIGR human gene index [6], STACK [7], the Merck gene index [8], and GeneNest [9]. The largest public, human gene index is the Unigene database with over 4 million sequences assembled into approximately 108,000 clusters (August 2003). The sequences represented in these databases have been extremely useful in applications that include constructing physical and sequence maps of the human genome [10–13], identifying genetic polymorphisms [14,15], annotating genomic sequences [16], finding new members of gene families [17], and locating exons and alternative splice sites [17–21]. As most tag collections are highly redundant, abundance levels for transcript tags have also provided comprehensive quantitative views of gene expression such that alterations in transcript levels associated with a variety of pathological conditions have been identified [22,23].

In this study, we have utilized a specialized prostate gene index termed the prostate expression database (PEDB) to characterize the repertoire of genes expressed in the normal and neoplastic human prostate through the analysis of 99,448 expressed sequence tags derived from 42 prostate cDNA libraries. Sequence assemblies and annotations identified 24,580 unique TUs of which more than 9,500 represent transcripts expressed in more than one prostate tissue source. Comparative analyses between transcriptomes derived from normal and neoplastic prostate tissues identified 75 genes with statistically-significant differential expression. Comparative analyses between the prostate and transcriptomes derived from other normal tissues identified more than 50 known and novel transcripts with expression enriched or restricted to the prostate. One uncharacterized TU predicted to exhibit prostate-restricted expression was cloned and prostate-specificity was verified by quantitative PCR. Further analysis of this TU revealed that it represented novel sequence extending a partial cDNA designated PCANAP5, a gene previously implicated in prostate carcinogenesis [24]. Together, these results define a subset of the TUs encoded in the human genome that are expressed in the prostate. This cohort can be used for computationally-based analyses and for designing customized microarrays comprising TUs of relevance for investigations of prostate gene expression involving development, hormone regulation, and pathology. The diversity of tissue sources used to generate the prostate transcriptome also allow for the determination and analysis of gene polymorphisms and alternative splicing events that may influence the biological consequences and complexity of gene expression.

MATERIALS AND METHODS

Analysis of Expressed Sequences in the Unigene Database

Analyses were performed on 56 prostate libraries archived in Unigene Build 153 (Supplemental Table S1). Several libraries were excluded from the analysis: ORESTES prostate libraries were excluded because preliminary studies revealed that several these libraries were not randomly sampled; NCI_CGAP_Sub8 and NCI_CGAP_Sub2 were excluded because they were mis-identified as prostate libraries in Unigene Build 153. Most of these libraries have been removed from later builds of the Unigene database.

Unigene clusters were searched for prostate-derived ESTs using a library identification number (LID) that associates each EST with a cDNA library and tissue of origin. Clusters containing prostate ESTs were retained and further divided into groups with or without ESTs derived from other tissues. ESTs found exclusively in libraries from prostate tissues were designated prostate-specific. To identify additional ESTs with prostate-enhanced, but not necessarily prostate-specific expression, we assigned each individual EST to a tissue type based upon the corresponding LID. Cell lines were included under their tissue of origin. To facilitate the

comparisons, we consolidated 362 tissue descriptions to 54 tissue types. Thirty-one tissue types were each represented by $\geq$10,000 ESTs and were used in this study (Supplemental Table S2). To determine relative EST abundance levels in each tissue, the number of ESTs within a cluster derived from a particular tissue type was normalized relative to the total number of ESTs for that tissue type. Unigene clusters that contained ESTs from prostate tissues that were expressed in no more than 5 out of 30 other tissues were chosen for further study. From this group, genes which had 3–5 fold higher levels of ESTs in the prostate relative to other tissues were considered prostate enriched.

Analysis of Expressed Sequences in the PEDB

The PEDB contains 119,855 raw ESTs with corresponding chromatogram files from libraries created in our laboratory [25,26] or obtained from the CGAP database (http://cgap.nci.nih.gov). A description of the libraries is available at http://www.pedb.org. A detailed description of the database and data analysis pipeline used to annotate the sequences has been described previously [26] and is also available at the PEDB web site. Briefly, repetitive DNA and vector sequences were masked and sequences that did not have 80/100 nucleotides with a Phred quality score greater than 20 were discarded. Phrap (P. Green, University of Washington, Seattle) was used to assemble contigs with the following settings: minmatch 60, minscore 90, and vector_bound 30. Contigs were used to query the Unigene, GenBank, and dbEST databases (downloaded August 2002) for sequence matches using BLAST (E^{-22}, Score: 200). Unigene annotations served as the primary annotation for each cluster. In the absence of a Unigene annotation, GenBank mRNA annotations took precedence over annotations from the dbEST database. In total, 24,580 clusters were assembled. All sequence information is available at http://www.pedb.org.

The statistical method of Audic and Claverie [27] was used to test for differential expression of EST clusters between normal prostate, represented by ESTs from microdissected normal prostate epithelium, and neoplastic prostate, represented by a pool of ESTs from PIN, primary adenocarcinoma and metastasis/xenograft libraries. Comparative analyses were also performed between the normal microdissected epithelium and each of the individual neoplastic prostate subcategories; PIN, primary adenocarcinoma, and metastasis/xenograft. The Audic and Claverie method takes into account the influence of random fluctuations and sampling size when calculating the probability of differential expression between tag libraries of different sizes. It performs well in comparisons with other tests [28], and has previously been used to identify gene expression changes in the prostate [29].

Cloning and Assembly of the cDNA Corresponding to Unigene Hs.163909

PCR primers were designed from GenBank sequences AA573109 (5′- AAGCTAGACAGGCAGCAGGACA-3′) and BI024803 (5′-CTCATTGCGGACTCCAAACA-3′). PCR was performed using Taq DNA polymerase (Invitrogen) and Human Prostate Marathon-ready cDNA (Clontech). PCR products were cloned into the pCR2.1-TOPO vector using the TOPO TA Cloning kit (Invitrogen). Further 5′ sequence was generated by PCR using primers designed from GenBank sequences XM_297687.2 (5′-GGACTCTACTGCCGAGACCA-3′) and AA573109 (5′-TATCCAGAAAAGTCTCCCGCC-3′). Further 3′ sequence was generated by PCR using primers designed from GenBank sequences AF109300 (5′-TGGGAACCGTGAAGGGTGTCCAGT-3′ and 5′-AACAGCAGCCAGCGGTTCACACAG-3′) and from sequence of 163,909 clones (5′-GCGCTGGGGCAAGTGGAACAAGTA-3′ and 5′-CCCACTTTGCCTCGACTGCCCTTT-3′).

Northern Analysis

A DNA insert derived from clone B12 representing Hs.163909 was labeled with α-32 P dCTP (Amersham) using the Prime-It RmT Random Primer Labeling Kit (Stratagene) and used to hybridize a Multiple Tissue Northern Blot II (Clontech) using ExpressHyb hybridization solution (Clontech). The blot was exposed to Kodak AR film at -80°C with intensifying screens.

Quantitative PCR

Expression analysis was performed on an ABI PRISM 7700 sequence detection system (Applied Biosystems) using SYBR Green PCR Master Mix (Applied Biosystems). First strand cDNA templates were generated with Superscript II reverse transcriptase (Invitrogen) from a tissue panel total RNAs (Clontech) primed with random hexamers. Primer pairs were designed using Primer Express software (Applied Biosystems) and are as follows: Hs.163909 (5′-AAGCTAGACAGGCAGCAGGACA-3′ and 5′-TACCTGGTCTACACACAGCCCA-3′), TRPM7 (5′-TTGTTTATGGACAGCAGTGGCT-3′ and 5′-CCAGTCTCGGAATGGTAAGGAA-3′), Sp0229 (5′-CCTCCATGCCTTCAAGAAACA-3′ and 5′-CAGGCAAAGGTGGAAGCTATTC-3′), and GAPDH (5′-TCTGCTCCTCCTGTTCGACAGT-3′ and 5′-ACGACCAAATCCGTTGACTCC-3′).

DNA Sequence Determination

DNA sequencing was performed using dye terminator sequencing reaction mix (Applied Biosystems) and analyzed on an ABI 3730 automated sequencer (Applied Biosystems). Sequence assembly and editing was performed using Sequencher software (Gene Codes Corp.).

RESULTS AND DISCUSSION

Diversity of the Prostate Transcriptome

To assess the diversity of gene expression in the human prostate, we first queried the Unigene database for prostate-derived ESTs. The Unigene database is the most comprehensively annotated public repository of expressed human sequences, providing annotations for roughly 102,000 putatively distinct TUs assembled from more than 3 million ESTs. Of these, 88,469 ESTs are derived from prostate sources and annotate to 16,652 distinct Unigene TUs. To enlarge the size of the Unigene-defined prostate transcriptome, we pooled the resources of the Unigene database with those of the PEDB (www.pedb.org). In addition to prostate ESTs obtained from public sequence repositories such as CGAP, the PEDB contains 13,022 high quality ESTs derived from cDNA libraries constructed in our laboratory [25,26,30]. A total of 117,434 ESTs were entered into the PEDB analysis process. Of these, 15,470 were excluded following stringent quality assessment that removed sequences comprising poor quality, repetitive elements, and cloning vector. A further 3,222 ESTs annotated to the mitochondrial genome and were not included in the assembly process. The remaining 99,448 ESTs were assembled into 24,580 clusters representing putatively distinct TUs comprising the prostate transcriptome. Based on sequence similarity as determined by the BLAST algorithm, approximately 15,000 TUs were assigned to previously identified Unigene clusters. The majority of these clusters (14,305) are represented by ESTs from Unigene prostate libraries; however, inclusion of additional sequences from the PEDB identified a further 900 TUs with Unigene annotations that are expressed in the prostate.

In addition to clusters with representation in the Unigene database, 9,425 clusters were identified in the PEDB that have no corresponding Unigene TU. Of these, 104 clusters annotate to cDNAs deposited in the GenBank database, and most of the remaining clusters have strong sequence similarity to ESTs in the dbEST database. Only 111 EST clusters did not have high quality sequence matches in any public database. These data suggest that the vast majority of the prostate EST clusters represent genuine transcripts and that at least 25,000 different transcripts are expressed in the human prostate. However, ESTs are subject to a variety of artifacts, the most prevalent of which is genomic DNA contamination of cDNA libraries. Thus a sequence match in dbEST does not necessarily correspond to an authentic transcript. Also, from a practical perspective, many of the clusters represented by only a single EST may correspond to rare transcripts that would be difficult to analyze using common expression quantitation methods such as microarray hybridization, quantitative PCR, or Northern analysis. Further studies focused on a cohort of 9,775 prostate TUs represented by ESTs derived from at least 2 prostate tissue sources.

Identification of Genes With Prostate Specific Expression

To identify genes that are uniquely expressed in the prostate, we queried the Unigene database for entries composed exclusively of ESTs with library identification tags (LIDs) from prostate tissue sources. There were 1,316 prostate-specific entries, but only 61 had ESTs from more than one prostate library (Supplemental Table S3). The most abundant prostate-specific cluster represents the *PRAC* gene, which encodes a prostate-specific small nuclear protein [31]. None of the remaining clusters have been biochemically verified to exhibit prostate-restricted expression. To choose a cluster for experimental validation, we mapped the sequence assemblies onto the draft human genome sequence (http://genome.ucsc.edu) to search for evidence of structural features suggestive of a transcribed gene. One cluster, Unigene Hs.163909, represents an uncharacterized TU that was found to partition into three exons when aligned with the genome sequence on chromosome 2q37. Hs.163909 is comprised of four ESTs derived from three prostate cDNA libraries. Northern blot analysis with a single EST clone from the cluster, identified two transcripts of 4.4 and 4.8 kb in RNA extracted from whole prostate tissue (Fig. 1A). Expression of Hs.163909 was not observed in seven other tissues assessed (spleen, thymus, testis, uterus, small intestine, colon, perhipheral blood leukocytes). Quantitative PCR was used to further examine the relative level of Hs.163909 mRNA expression in 19 tissues. Hs.163909 was highly expressed in whole prostate with a very low level detected in colon, small intestine, and stomach samples (Fig. 1B).

All of the ESTs in the Hs.163909 cluster map within a 14 kb region of the human draft genome sequence bounded by the protein phosphatase 1 regulatory subunit number 7 (PPP1R7) locus and the prostate cancer associated protein 5 (PCANAP5) locus. The proximity of Hs.163909 to a gene implicated in prostate cancer [24], and the identification of at least four other

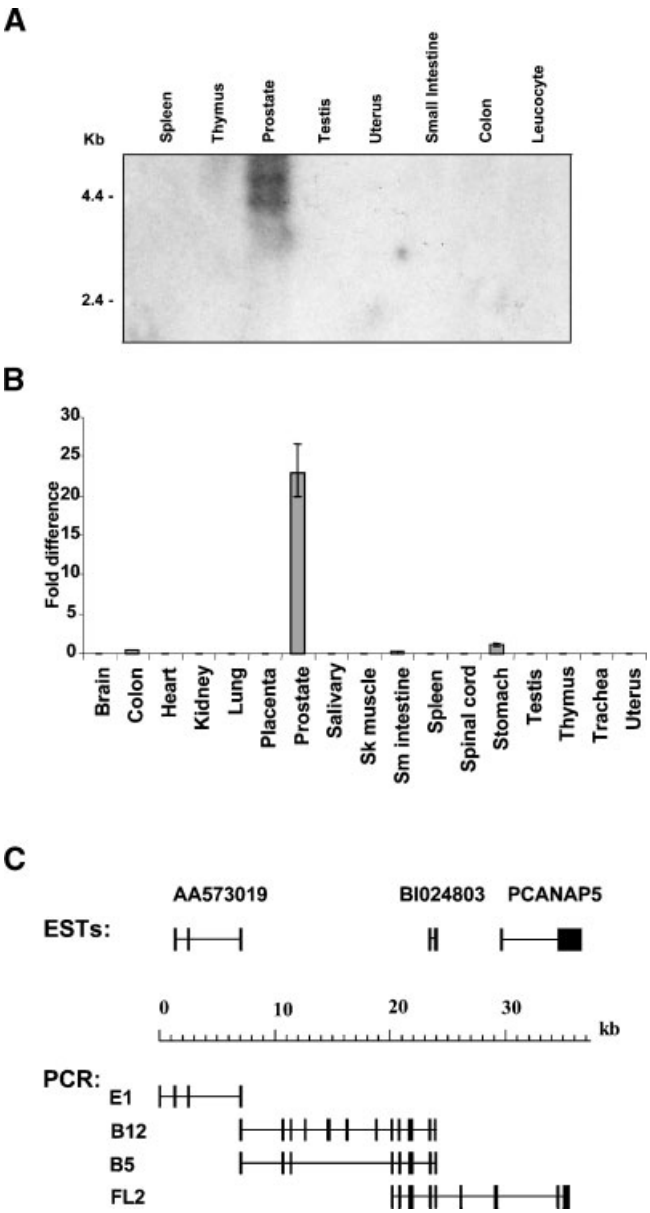

Fig. 1. Analysis of prostate-specific TU Hs.163909. **A**: Northern blot of multiple tissues probed with an Hs.163909 cDNA. **B**: Expression of Hs.163909/PCANAP5 in 17 tissues measured by QPCR. Fold differences were calculated with respect to GAPDH transcript amplification, then normalized relative to expression levels in stomach tissue (1 U). Sk muscle, skeletal muscle; Sm intestine, small intestine. **C**: cDNA clones aligned with genomic DNA. Vertical bars represent exons, horizontal lines introns. ESTs are from the Unigene database. PCR products are sequenced cDNA clones created using RT-PCR and gene-specific primers. Clones E1 and FL2 were derived from prostate mRNA; clones B12, B5, from colon mRNA. (Three additional ESTs, and PCR products, not shown.)

prostate ESTs that map within the same 14 kbp region (but not included in Unigene Build 153) led us to further characterize Hs.163909. Sequencing of RT-PCR-derived clones created using primers designed from Hs.163909 ESTs, PCANAP5 (Hs.35165), and prostate ESTs not in Unigene, reveal that all of the clones represent a single transcription unit with multiple potential splice variants (Fig. 1C). Nineteen separate exons were detected amongst cDNA clones isolated by RT-PCR from prostate and colon RNA. The conceptually translated proteins from multiple possible alternative splice forms of Hs.163909 are not similar to any other known human protein. A single spliced mouse EST that maps to a chromosomal position adjacent to the mouse ortholog of PPP1R7 (BB649330) shares amino acid similarity with PCANAP5 suggesting that this gene is conserved in mammalian species.

The *PCANAP5* gene (formerly designated Implicated in Prostate Cancer 5) was originally identified by a statistical analysis of 522 EST libraries in a proprietary database which indicated that the pattern of PCANAP5 expression was more similar to prostate-specific genes associated with cancer than other genes [24]. To verify this conclusion, we screened 10 matched cancer and normal prostate samples using QPCR, but found no evidence for consistent up- or down-regulation of PCANAP5 expression in prostate cancer (data not shown). Since PSA is androgen regulated, we further speculated that the statistical association between PCANAP5 and PSA was a consequence of androgen regulation; however, QPCR of RNA from androgen-stimulated and-starved LNCaP cells did not reveal any evidence of androgen regulation (data not shown; see method in [32]). Thus, the role of PCANAP5 in prostate development, physiology, and disease remains unknown.

Identification of Genes With Enhanced Expression in the Prostate

The majority of transcribed genes are not absolutely restricted in expression to a single tissue. To search for genes that are expressed at relatively higher levels in the prostate compared to other tissues, we applied two strategies. First, the Unigene database was searched for clusters that contained a high proportion of prostate ESTs relative to the sum of ESTs derived from all other tissues. In the second strategy, individual clusters were partitioned by tissue type and the number of ESTs for each tissue was statistically compared to the number of prostate ESTs.

Eleven Unigene clusters were comprised of 5 fold more prostate ESTs than the combined total of ESTs derived from other tissues (Supplemental Table S4). The four clusters with the highest number of ESTs represented genes previously reported to exhibit profiles highly restricted to the prostate gland: β-microseminoprotein, seminogelin 1, seminogelin 2, and transglutaminase 4. The other seven clusters have not been studied in detail. Interestingly, none of the ESTs

comprising the prostate-enhanced transcript represented by Unigene Hs.372058 annotate to the draft human genome sequence (February 2003), demonstrating that even with careful curation, the current genome assembly annotations are incomplete. The gene encoded by Hs.360089 is predicted to be expressed only in the prostate and the lung. Aligning this sequence with the human genome assembly (http://genome.ucsc.edu) indicated splicing into at least two exons. Hence it also represents a good candidate for further study.

A more thorough method of searching for genes preferentially expressed in the prostate resulted from dividing Unigene ESTs into subsets based on tissue type and then comparing the subsets for tissue distribution. We compressed 362 different Unigene tissue descriptions into 54 general tissue categories and determined the number of Unigene ESTs in each category. Thirty-one tissue categories each with at least 10,000 ESTs were used for further analysis (Supplemental Table S2). Ninety-four Unigene entries with prostate-derived ESTs were present in at least 2 prostate libraries and not in the 30 other tissues (data not shown). As describe above, 61 were exclusively found in prostate libraries, while 33 were also expressed in tissue types with libraries that did not have the 10,000 ESTs we used as a cutoff for statistical analysis of differential expression.

Using the method of Audic and Claverie [27], we identified 45 genes that are differentially expressed in the prostate relative to at least 25 of the 30 tissue types comprising the multi-tissue analysis pool ($P > 0.99$; Supplemental Table S5). Sixteen of the genes encoded ribosomal proteins. Ten of the non-ribosomal TUs were differentially expressed in prostate relative to all 30 non-prostate tissues. The most abundant differentially expressed gene is *RBBP7*; however, the vast majority of RBBP7 ESTs (97%) originated from a single library (NIH_MGC_60), suggesting a sampling bias. Of the remaining 29 non-ribosomal genes, 8 are expressed at least 5 fold higher in the prostate relative to any other tissue: SEMG1, SEMG2, KLK2, ACPP, TGM4, NKX3A, TRGγ, and MSMB; two: TRPM7, KLK3, are expressed at least 5 fold higher in the prostate relative to all but one other tissue (tonsil and breast, respectively); one, FLJ21128, is expressed over 4 fold higher in the prostate; and 2 others; ARSDR1 and ACTG2, are expressed between 3–4 fold higher in prostate than in any other tissue (Table I).

Most of the genes predicted to be prostate enriched have previously been experimentally verified. These include transglutaminase 4 (TGM4), β-microseminoprotein (MSMB), Seminogelin 1 and 2 (SEMG1, 2), retinol dehydrogenase 11 (ARSDR1), and the androgen-regulated genes kallikrein 2 and kallikrein 3 (KLK2, KLK3), prostate acid phosphatase, (ACPP), NKX3 transcription homolog A (NKX3A), and transmembrane protease serine 2 (TMPRSS2) [33–40]. The TRPM7 transcript encodes transient receptor potential cation channel M7, a protein belonging to a family of calcium-selective channel proteins [41,42] that has not been extensively studied in the prostate. To verify the tissue distribution of TRPM7 expression we performed

TABLE I. Genes With High Expression in the Prostate Relative to Other Tissues

Gene	Description	Tissues expressing the gene[a]	Minimum fold difference[b]	Prostate ESTs[c]	Unigene ID
MSMB	Microseminoprotein-beta	7	>5	536	Hs.183752
KLK3	Kallikrein 3, (prostate specific antigen)	9	>5	500	Hs.171995
ACPP	Acid phosphatase, prostate	12	>5	344	Hs.1852
SEMG1	Semenogelin 1	2	>5	225	Hs.1968
KLK2	Kallikrein 2, prostatic	6	>5	153	Hs.181350
SEMG2	Semenogelin II	2	>5	76	Hs.180016
TGM4	Transglutaminase 4 (prostate)	2	>5	50	Hs.2387
TRG	T cell receptor gamma locus	14	>5	48	Hs.112259
NKX3A	NK3 transcription factor homolog A (*Drosophila*)	11	>5	45	Hs.55999
TRPM7	Transient receptor potential cation channel, subfamily M, member 7	25	>5	63	Hs.267914
ARSDR1	Androgen-regulated short-chain dehydrogenase/reductase 1	28	3–4	118	Hs.179817
ACTG2	Actin, gamma 2, smooth muscle, enteric	22	3–4	87	Hs.78045
FLJ21128	Hypothetical protein FLJ21128	25	3–4	52	Hs.288389

[a]Tissues are listed in the supplementary table, Table S2.
[b]Minimum ratio (prostate Ests/total prostate ESTs)/(tissue Ests/total tissue ESTs) across 29/30 tissues.
[c]Total prostate ESTs: see Table S2.

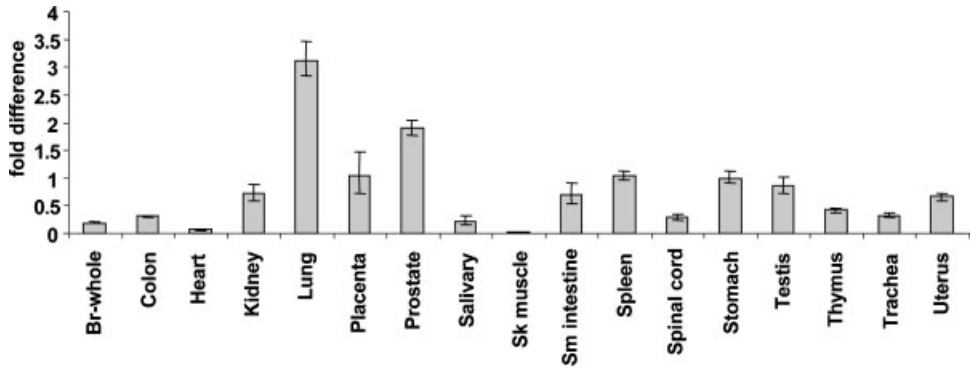

Fig. 2. Enriched expression of TRPM7 (transient receptor cation channel, subfamily M, member 7) in the prostate and lung determined by QPCR. Fold differences were calculated with respect to GAPDH transcript amplification, then normalized relative to fold difference in stomach tissue. Sk muscle, skeletal muscle; Sm intestine, small intestine.

QPCR with RNA from 20 tissues and 2 prostate-derived cell lines. TRPM7 was most highly expressed in lung, prostate, and fetal brain tissues and was expressed at least 2 fold higher in the prostate than in 11 of 16 other tissues analyzed (Fig. 2). Thus, while TRPM7 is prostate-enriched, its expression in the prostate relative to other tissues is not as robust as predicted by EST analysis. Since Cat-L, another member of the same ion channel family, is differentially expressed in primary adenocarcinomas of the prostate [43], we evaluated 10 matched samples of normal and neoplastic prostate tissue for TRPM7 dysregulation, but found no evidence of differential expression (data not shown). We also did not observe androgen-mediated regulation of TRPM7 expression in LNCaP cells (data not shown).

Novel Transcripts Expressed in the Prostate

In addition to the genes expressed in the prostate that are represented in the Unigene database, we identified 475 prostate TUs without Unigene annotations that contained ESTs from at least 2 independently derived prostate cDNA libraries. Of these, 48 TUs are represented by 5 or more ESTs. When these TUs were screened against the database of ESTs using the BLAST program, 18 of the clusters had strong sequence similarity to ESTs derived only from prostate cDNA libraries, suggesting that they may represent prostate-specific genes or alternatively spliced exons (Supplemental Table S6). A complementary search strategy was also employed that relied on the identification of transcript splicing to provide evidence that a novel sequence actually represented a genuine transcript. We searched the human genome draft sequence (http://genome.ucsc.org) with 1,117 TUs from PEDB without Unigene annotations that were comprised of at least 2 ESTs, but did not require that the ESTs originate from different libraries. Ten TUs were found to partition into more than one exon (Supplemental Table S7). Only four of these TUs (Sp0423, Sp0116, Sp0531, and Sp0229) did not have sequence similarity to ESTs derived from other tissues. One cluster, Sp0151 was comprised of ESTs from only prostate and breast, and maps within an intron of the *TRG* gene.

The Sp0229 cluster was chosen for further study because it was not found in the Unigene database and was composed solely of ESTs from the PEDB. Aligning the Sp0229 cluster with the human genome draft sequence revealed that it is located at position 12q21.31 and is transcribed from the opposite DNA strand as the protein tyrosine phosphatase gene *PPFIA2*. An additional prostate-derived sequence, PEDB: 137093, maps within the first intron defined by all of the spliced clones, suggesting that it is also part of the same cluster. Taken together, 4–5 ESTs identify a single TU with at least six exons that span a 56 kb region of genomic DNA (Fig. 3A). Gene prediction software does not identify the *Sp0229* gene (http://genome.ucsc.edu) and an open reading frame is not apparent. QPCR (Fig. 3B) and Northern analysis (data not shown) identified Sp0229 transcripts in several tissues including those of neuronal (brain and spinal chord), lung, stomach, testis, and prostate origin, and essentially no expression in 11 of 14 other tissues. While these findings demonstrate the utility of combining a variety of EST and genomic resources to identify novel transcripts that are expressed in the prostate, they also illustrate the limitations of database analyses and reinforce the need for biological experiments to verify results.

Transcripts With Altered Expression in Prostate Carcinoma

A major goal of prostate cancer research is to identify consistent molecular alterations between normal and neoplastic tissues such that they may exploited for understanding the process of carcinogenesis, and

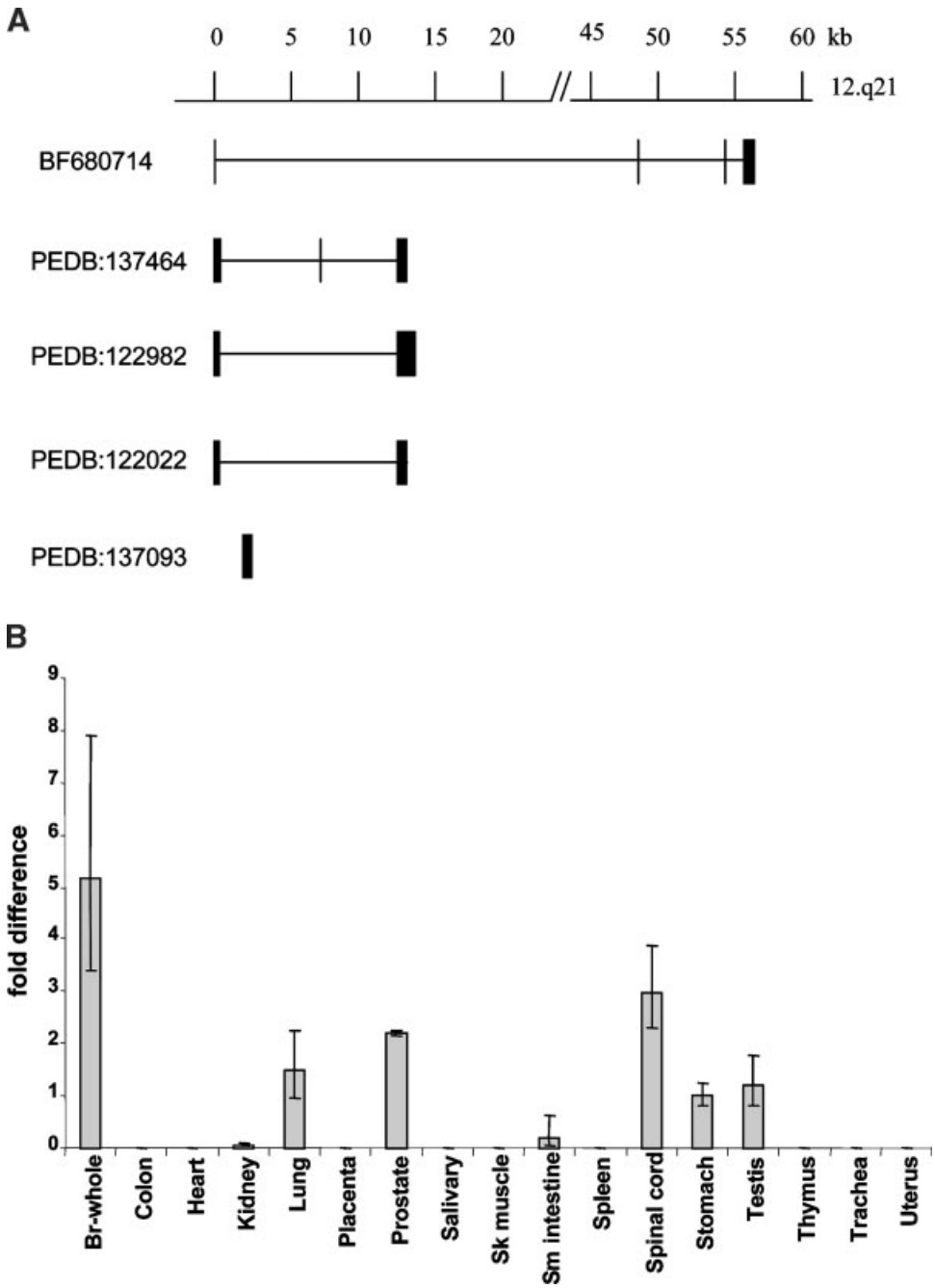

Fig. 3. A novel transcription unit, Sp0229, identified from EST clusters in PEDB. **A**: Alignment of cDNAs with genomic DNA. Vertical bars represent exons, horizontal lines introns. **B**: Sp0229 expression in a variety of tissues, as determined by QPCR. Relative fold differences were calculated with respect to GAPDH transcript amplification, then normalized relative to fold difference in stomach tissue. Sk muscle, skeletal muscle; Sm intestine, small intestine.

Fig. 4. Patterns of gene expression calculated from EST frequencies in cDNA libraries. Relative expression levels are shown as fold-changes between normal prostate epithelium, PIN, primary prostate adenocarcinoma, and prostate cancer metastases and xenografts. **A**: Genes with over 5,000 ESTs in the dataset. **B**: Genes with fewer than 5,000 ESTs in the dataset. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

Symbol	Unigene	Description
----	----	pedb:15888
PRKAR1B	Hs.1519	cAMP-dependent kinase
MGAT4B	Hs.4867	β-1,4-N-acetylglucosaminyltransferase
MSMB	Hs.183752	β-microseminoprotein
KLK3	Hs.171995	kallikrein 3
ACPP	Hs.1852	prostate acid phosphatase
----	----	pedb:Sp3
----	----	pedb:16830
RPL37A	Hs.296290	ribosomal protein L37a
RPS17	Hs.5174	ribosomal protein S17
RPL31	Hs.184041	ribosomal protein L31
----	Hs.356428	mRNA expressed only in the placenta
RPS18	Hs.275865	ribosomal protein S18
RPL7	Hs.153	ribosomal protein L7
----	Hs.301809	EST

Symbol	Unigene	Description
----	Hs.332048	EST
RPL38	Hs.2017	ribosomal protein L38
----	Hs.301104	EST
EEF2	Hs.75309	translation elongation factor 2
FLJ32334	Hs.374785	hypothetical protein FLJ32334
ARHU	Hs.20252	ras homolog gene family member U
RPL7	Hs.356593	ribosomal protein L7
----	----	pedb:22335
HSGP25L2G	Hs.279929	gp25L2 protein
ALCAM	Hs.10247	activated leucocyte cell adhesion molecule
FABP5	Hs.153179	fatty acid binding protein 5
----	----	pedb:22482
FKBP3	Hs.350402	FK506 binding protein 3
PABPC1	Hs.172182	poly(A) binding protein
ACTG2	Hs.78045	actin gamma 2
ACTA2	Hs.195851	actin alpha 2
MYL6	Hs.77385	myosin light polypeptide 6
FTH1	Hs.62954	ferritin heavy polypeptide 1
TAGLN	Hs.75777	transgelin
FLJ21174	Hs.194329	hypothetical protein FLJ21174
UQCRC2	Hs.173554	ubiquinol-cytochrome c reductase core protein 2
PLXNB2	Hs.3989	plexin B2
CKB	Hs.173724	creatine kinase, brain
LAMB2	Hs.90291	beta 2 laminin
RPL27	Hs.111611	ribosomal protein L27
RPL23A	Hs.350046	ribosomal protein L23a
RPLP1	Hs.177592	ribosomal protein, large, P1
SERPINA3	Hs.234726	serine proteinase inhibitor
H3F3A	Hs.181307	H3 histone, family 3A
UQCRH	Hs.73818	ubiquinol-cytochrome c reductase
PSMA6	Hs.74077	proteasome subunit
NACA	Hs.32916	nascent polypeptide associated polypeptide
RPL19	Hs.252723	ribosomal protein L19
EIF3S8	Hs.4835	translation initiation factor 3
GNB2L1	Hs.5662	GNBP beta polypeptide 2-like 1
ACTG1	Hs.14376	actin gamma 1
XBP1	Hs.149923	X-box binding protein
H19	Hs.334822	H19, maternally expressed
SAT	Hs.28491	Spermidine/spermine N1-acetyltransferase
TRG	Hs.112259	T cell receptor gamma locus
VPS35	Hs.264190	vacuolar protein sorting 35 (yeast)
LOC54543	Hs.112318	6.2 kd protein

Fig. 4.

provide diagnostic markers and therapeutic targets. DNA microarrays, differential display, and electronic profiling have each been used to discover gene expression changes that associate with prostate carcinoma [24,44–59]. As the cDNA libraries contributing ESTs to Unigene and PEDB were derived from a spectrum of normal and neoplastic prostate tissues, we employed a statistical method [27] to identify differentially-expressed transcripts between normal prostate epithelium and prostate cancer. For the initial comparison, all ESTs derived from normal prostate tissues were combined into one group, and all ESTs derived from neoplastic prostate sources were combined into another. In total, 8,794 genes were tested for differential expression between these two datasets. The abundance levels of 75 genes were significantly different between the two samples ($P > 0.99$; Supplemental Tables S8 and S9). Of these, 5 genes were up-regulated in cancer relative to normal prostate epithelium and 70 were down-regulated. When prostate cancers were further sub-divided into prostate intraepithelial neoplasia (PIN), primary adenocarcinoma and metastasis/xenograft categories, 37 genes were found to be upregulated in at least one sub-category relative to normal prostate epithelium, and 20 were downregulated (Fig. 4).

We compared these results with a published list of 46 genes reported to be differentially expressed in prostate carcinoma identified using a different comparative analysis algorithm also based on database sequence tags [51]. Eleven genes were identified in both studies. Of these, five genes were determined to be upregulated in cancer: ras homolog gene family member U (*ARHU*); ribosomal protein L19 (*RP19*); eukaryotic translation elongation factor 2 (*EEF2*); fatty acid binding protein 5 (*FABP5*); 6.2 kd protein, *LOC54543*, and six genes were downregulated in cancer: brain creatine kinase (*CKB*); fibulin 5 (*FBLN5*); hypothetical protein FLJ21174 (*FLJ21174*); ubiquinol-cytochrome c reductase core protein 2 (*UQCRC2*); 9.5 kDa ubiquinone binding protein (*QP-C*); and neuroblastoma suppression of tumorogenicity 1 (*NBL1*). The differential expression of NBL1 was confirmed by laboratory experimentation [51].

Several of the genes that were determined to exhibit altered expression levels in prostate carcinoma through statistical analyses of the PEDB dataset have been independently reported to have altered gene expression in at least a subset of prostate cancers using biochemical approaches. These include ACTG2, ALCAM, TRG, SEMG2, MSMB, SAT, TAGLN, and NBL1 [47–49,51,54,60]. Thus, statistical methods can in principle identify mis-regulated genes. To identify additional markers for prostate cancer, we concentrated on genes that had high probabilities of dysregulation both when the cancer samples were grouped together, and when at least one cancer sub-category was statistically different from normal epithelium. These two different methods of stratification shared 15 genes. We analyzed the expression of four: UQCRC2, FLJ21174, LAMB2, and H19 in six matched pairs of normal and cancerous prostate tissues using QPCR and did not observe consistent expression level differences (data not shown). Thus, while statistical methods can identify genes with altered expression in prostate cancer, the inherent heterogeneity of prostate cancer mandates verification using biochemical methods in multiple cancer samples. Due to the labor involved in library construction and DNA sequencing, most sequence-based expression studies such as those enumerating ESTs or employing serial analysis of gene expression (SAGE) methods actually compare relatively few different tissues [61–63]. The prostate cancer sample pool used in this analysis comprises only 15 carcinoma samples. The results of cDNA microarray experiments measuring transcript alterations in scores of cancer cases and tissue microarray studies evaluating protein expression in hundreds of histological samples indicate that large cohorts of individual cancers must be compared in order to assess the considerable variability of gene expression in prostate neoplasia [47,64]. Pooling multiple different samples may average out expression differences in all but the most robustly dysregulated genes (Table II).

CONCLUSIONS

The analysis presented here indicates that the human prostate gland expresses a diverse repertoire of genes that reflect a functionally complex organ with highly specialized secretory function. We currently estimate the prostate transcriptome to comprise approximately 25,000 distinct TUs though a more conservative estimate requiring a given transcript to be present in more than one prostate tissue source comprises approximately 10,000 different expressed genes. A list of these genes is available at http://www.pedb.org. However, the gene cohorts that currently define the prostate transcriptome will likely both expand and contract as human genome assemblies and annotations continue to be refined. Some prostate TUs currently annotated as independent gene products will condense as true exon relationships are defined as with Unigene clusters Hs.163909 and Hs.35165 that both represent different regions of the *PCANAP5* gene. As additional ESTs are added to the Unigene and PEDB databases, sequences that currently cluster together based on small regions of shared homology may be re-partitioned into homologous but distinct TUs. Genes not currently included in the prostate transcriptome may be expressed only during

TABLE II. Differentially Expressed Genes in Prostate Cancer

Gene	UniGene ID	PEDB ID	Description	ESTs per million	
				Normal	Cancer
MSMB	Hs.183752	4175	Beta-microseminoprotein	14,896	7,693
ACTG2	Hs.78045	12383	Actin, gamma 2, smooth muscle, enteric	2,888	1,225
MYL6	Hs.77385	1411	Myosin, light polypeptide 6	2,736	1,176
ACTA2	Hs.195851	2345	Actin, alpha 2, smooth muscle, aorta	1,824	539
CKB	Hs.173724	14755	Brain creatine kinase	1,368	196
PLXNB2	Hs.3989	15156	Plexin B2	912	—
UQCRC2	Hs.173554	2845	Ubiquinol-cytochrome c reductase core protein II	912	49
FLJ21174	Hs.194329	15111	Hypothetical protein FLJ21174	912	49
LAMB2	Hs.90291	7288	Laminin, beta 2	912	98
—	—	16631	—	456	—
RPS17	Hs.5174	10537	Ribosomal protein S17	2,584	5,439
GNB2L1	Hs.5662	7185	G protein, beta polypeptide 2-like 1	1,520	3,430
None	Hs.301809	3366	EST	1,064	3,724
EIF3S8	Hs.4835	9425	Eukaryotic translation initiation factor 3, subunit 8 (110 kD)	—	931
H19	Hs.334822	4936	Imprinted maternally expressed untranslated Mrna	—	882

specific stages of development, or may require the influence of endocrine and paracrine factors such as androgenic hormones. The continued analysis of prostate gene expression both through sequence-based methods and high-throughput microarray and proteomics approaches will provide a more complete picture of how genes and their products interact to form the developmental, signaling, and metabolic pathways that are the basis for normal and pathological processes involving the prostate.

REFERENCES

1. Boguski MS, Lowe TM, Tolstoshev CM. dbEST-database for 'expressed sequence tags.' Nat Genet 1993;4:332–333.
2. Schaefer C, Grouse L, Buetow K, Strausberg RL. A new cancer genome anatomy web resource for the community. Cancer J 2001;7:52–60.
3. Lash AE, Tolstoshev CM, Wagner L, Schuler GD, Strausberg RL, Riggins GJ, Altschul SF. SAGEmap: A public gene expression resource. Genome Res 2000;10:1051–1060.
4. Okazaki Y, Furuno M, Kasukawa T, et al. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. Nature 2002;420:563–573.
5. Schuler GD. Pieces of the puzzle: Expressed sequence tags and the catalog of human genes. J Mol Med 1997;75:694–698.
6. Quackenbush J, Cho J, Lee D, Liang F, Holt I, Karamycheva S, Parvizi B, Pertea G, Sultana R, White J. The *TIGR* gene indices: Analysis of gene transcript sequences in highly sampled eukaryotic species. Nucleic Acids Res 1999;29:159–164.
7. Miller RT, Christoffels AG, Gopalakrishnan C, Burke J, Ptitsdyn AA, Broveak TR, Hide WA. A comprehensive approach to clustering of the expressed human gene sequence: The sequence tag alignment and consensus knowledge base. Genome Res 1999;9:1143–1155.
8. Eckman BA, Aaronson JS, Borowski JA, Bailey WJ, Elliston KO, Williamson AR, Blevins RA. The Merck gene index browser: An extensible data integration system for gene finding, gene characterization and EST data mining. Bioinformatics 1998;14: 2–13.
9. Krause A, Haas SA, Coward E, Vignron M. SYSTERS, GeneNest, SpliceNest: Exploring sequence space from genome to protein. Nucleic Acids Res 2002;30:299–300.
10. Liang F, Holt I, Pertea G, Karamycheva S, Salzberg SL, Quackenbush J. Gene index analysis of the human genome estimates approximately 120,000 genes. Nat Genet 2000;25:239–240.
11. Hudson TJ, Stein LD, Gerety SS, Ma J, Castle AB, Silva J, Slonim DK, Baptista R, Kruglyak L, Xu SH. An STS-based map of the human genome. Science 1995;274:540–546.
12. Schuler GD, Boguski MS, Stewart EA, Stein LD, Gyapay G, Rice K, White RE, Rodriguez-Tome P, Aggarwal A, Bajorek E, et al. A gene map of the human genome. Science 1996;274:540–546.
13. Deloukas P, Schuler GD, Gyapay G, Beasley EM, Soderlund C, Rodriguez-Tome P, Hui L, Matise TC, McKusick KB, Beckmann JS, et al. A physical map of 30,000 human genes. Science 1998, 282:744–746.
14. Garg K, Green P, Nickerson DA. Identification of candidate coding region single nucleotide polymorphisms in 165 human genes using assembled expressed sequence tags. Genome Res 1999;9:1087–1092.
15. Buetow KH, Edmonson MN, Cassidy AB. Reliable identification of large numbers of candidate SNPs from public EST data. Nat Genet 1999;21:323–325.
16. Waterston R, Martin C, Craxton M, Huynh C, Coulson A, Hillier L, Durbin R, Green P, Halloran N, et al. A survey of expressed genes in *Caenorhabditis elegans*. Nat Genet 1992;1:114–123.
17. McCombie WR, Adams MD, Kelley JM, FitzGerald MG, Utterback TR, Khan M, Dubnick M, Kerlavage AR, Venter JC, Fields C. *Caenorhabditis elegans* expressed sequence tags identify gene families and potential disease gene homologues. Nat Genet 1992; 1:124–131.

18. Brett D, Hanke J, Lehmann G, Haase S, Delbruck S, Krueger S, Reich J, Bork P. EST comparison indicates 38% of human mRNAs contain possible alternative splice forms. FEBS Lett 2000;474: 83–86.
19. Mironov AA, Fickett JW, Gelfland MS. Frequent alternative splicing of human genes. Genome Res 1999;9:1288–1293.
20. Kan Z, Rouchka EC, Gish WR, States DJ. Gene structure prediction and alternative splicing analysis using genomically aligned ESTs. Genome Res 2001;11:889–900.
21. Kan Z, States D, Gish W. Selecting for functional alternative splices in ESTs. Genome Res 2002;12:1837–1845.
22. Adams MD, Kerlavage AR, Fleischmann RD, Fuldner RA, Bult CJ, Lee NH, Kirkness EF, Weinstock KG, Gocayne JD, White O, et al. Initial assessment of human diversity and expression patterns based upon 83 million nucleotides of cDNA sequence. Nature 1995;377:3–174.
23. Schmitt AO, Specht T, Beckmann G, Dahl E, Pilarsky CP, Hinzmann B, Rosenthal A. Exhaustive mining of EST libraries for genes differentially expressed in normal and tumour tissues. Nucleic Acids Res 1999;21:4251–4260.
24. Walker MG, Volkmuth W, Sprinzak E, Hodgson D, Klingler T. Prediction of gene function by genome-scale expression analysis: Prostate cancer-associated genes. Genome Res 1999;9:1198–1203.
25. Nelson PS, Ng WL, Schummer M, True LD, Liu AY, Bumgarner RE, Ferguson C, Dimak A, Hood L. An expressed-sequence-tag database of the human prostate: Sequence analysis of 1,168 cDNA clones. Genomics 1998;47:12–25.
26. Hawkins V, Doll D, Bumgarner R, Smith T, Abajian C, Hood L, Nelson PS. PEDB: The prostate expression database. Nucleic Acids Res 1999;27:204–208.
27. Audic S, Claverie JM. The significance of digital gene expression profiles. Genome Res 1995, 7:986–995.
28. Ruomualdi C, Bortoluzzi S, Danieli GA. Detecting differentially expressed genes in multiple tag sampling experiments: Comparative evaluation of statistical test. Hum Mol Genet 2001;19:2133–2141.
29. Clegg N, Eroglu B, Ferguson C, Arnold H, Moorman A, Nelson PS. Digital expression profiles of the prostate androgen-response program. J Steroid Biochem 2002;80:13–23.
30. Nelson PS, Pritchard C, Abbott D, Clegg N. The human (PEDB) and mouse (mPEDB) prostate expression databases. Nucleic Acids Res 2002;30:218–220.
31. Liu XF, Olsson P, Wolfgang CD, Bera TK, Duray P, Lee B, Pastan I. PRAC: A novel small nuclear protein that is specifically expressed in human prostate and colon. Prostate 2001;47: 125–131.
32. Nelson PS, Clegg N, Arnold H, Ferguson C, Bonham M, White J, Hood L, Lin B. The program of androgen-responsive genes in neoplastic prostate epithelium. Natl Acad Sci USA 2002;99: 11890–11895.
33. Hara M, Kimura H. Two prostate-specific antigens, gamma-seminoprotein and beta-microseminoprotein. Prostate 1988;12: 29–38.
34. Lilja H, Abrahamsson PA. Three predominant proteins secreted by the human prostate gland. Prostate 1998;12:29–38.
35. Riegman PH, Vliestra RJ, van der Korput HA, Romijn JC, Trapman J. Identification and androgen regulation of two major human glandular kallikrein-1 (hGK-1) mRNA species. Mol Cell Endocrinol 1991;76:181–190.
36. An G, Meka CS, Bright SP, Veltri RW. Human prostate-specific transglutaminase gene: Promoter cloning, tissue-specific expression, and down regulation in metastatic prostate cancer. Urology 1999;54:1105–1111.
37. Lundwall A, Bjartell A, Olsson AY, Malm J. Semenogelin 1 and 2, the predominant human seminal plasma proteins, are also expressed in non-genital tissues. Mol Hum Reprod 2002;8:805–810.
38. He WW, Sciavolino PJ, Wing J, Augustus M, Hudson P, Meissner PS, Curtis RT, Shell BK, Bostwick DG, Tindall DJ, Gelmann EP, Abate-Shen C, Carter KC. A novel human prostate-specific, androgen-regulated homeobox gene (*NKX3.1*) that maps to 8p21, a region frequently deleted in prostate cancer. Genomics 1997;43:69–77.
39. Lin B, Ferguson C, White J, Wang S, Vessella R, True LD, Hood L, Nelson PS. Prostate-localized and androgen-regulated expression of the membrane-bound serine protease TMPRSS2. Cancer Res 1999;59:4180–4184.
40. Lin B, White JT, ferguson C, Wang S, Vessella R, Bumgarner R, True LD, Hood L, Nelson PS. Prostate short-chain dehydrogenase reductase 1 (PSDR1): A new member of the short-chain steroid dehydrogenase/reductase family highly expressed in normal and neoplastic prostate epithelium. Cancer Res 2001;61:1611–1618.
41. Runnels LW, Yue L, Clapham DE. TRP-PLIK, a bifunctional protein with kinase and ion channel activities. Science 2001;291: 1043–1047.
42. Nadler MJS, Hermsura MC, Inabe K, Perraud A-L, Zhu Q, Stokes AJ, Kurosaki T, Kinet J-P, Penner R, scharenberg AM, Fleig A. LTRPC7 is a Mg.ATP-regulated divalent cation channel required for cell viability. Nature 2001;411:590–595.
43. Wissenbach U, Niemeyer BA, Fixemer T, Schneiderwind A, Trost C, Cavalie A, Reus K, Meese E, Bonkhoff H, Flockerzi V. Expression of CaT-like, a novel calcium-selective channel correlated with the malignancy of prostate cancer. J Biol Chem 2001;276:19461–19468.
44. Huang GM, Ng W, Farkas J, He L, Liang HA, Gordon D, Yu J, Hood L. Prostate cancer expression profiling by cDNA sequencing analysis. Genomics 1998;59:178–186.
45. Scheurle D, DeYoung P, Binninger DM, Page H, Jajanzeb M, Narayanan R. Cancer gene discovery using digital differential expression. Cancer Res 2000;60:4037–4043.
46. Luo J, Duggan DJ, Chen Y, Sauvageot J, Ewing CM, Bittner ML, Trent JM, Isaacs WB. Human prostate cancer and benign prostatic hyperplasia: Molecular dissection by gene expression profiling. Cancer Res 2001;61:4683–4688.
47. Dhanasekaran SM, Barrettte TR, Ghosh D, Shah R, Varambally S, Kurachi K, Pienta KJ, Rubin MA, Chinnaiyan AM. Delineation of prognostic biomarkers in prostate cancer. Nature 2001;412:822–826.
48. Magee JA, Araki T, Patil S, Ehrig T, true L, Humphrey PA, Catalona WJ, Watson MA, Milbrandt J. Expression profiling reveals hepsin overexpression in prostate cancer. Cancer Res 2001;61:5692–5696.
49. Welsh JB, Sapinosa LM, Su AI, Kern SG, Wang-Rodriguez J, Moskaluk CA, Frierson HF Jr., Hampton GM. Analysis of gene expression identifies candidate markers and pharmacological targets in prostate cancer. Cancer Res 2001;61:5974–5978.
50. Olsson P, Bera TK, Essand M, Kumar V, Duray P, Vincent J, Lee B, Pastan I. *GDEP*, a new gene differentially expressed in normal prostate and prostate cancer. Prostate 2001;48:231–241.
51. Asmann YW, Kosari F, Wang K, Cheville JC, Vasmatzis G. Identification of differentially expressed genes in normal and malignant prostate by electronic profiling of expressed sequence tags. Cancer Res 2002;62:3308–3314.

52. DePrimo SE, Diehn M, Nelson JB, Reiter RE, Matese J, Fero M, Tibshirani R, Brown PO, Brooks JD. Transcriptional programs activated by exposure of human prostate cancer cells to androgen. Genome Biol 2002;3:Research 0032.

53. Segawa T, Nau ME, Xu LL, Chilukuri RN, Makarem M, Zhang W, Petrovics G, sesterhenn IA, McLeod DG, Moul JW, Vahey M, Srivastava S. Androgen-induced expression of endoplasmic reticulum (ER) stress response genes in prostate cancer cells. Oncogene 2002;21:8749–8758.

54. Luo J, Zha S, Gage WR, Dunn TA, Hicks JL, Bennett CJ, Ewing CM, Platz EA, Ferdinandusse S, Wanders RJ, Trent JM, Isaacs WB, De Marzo AM. Alpha-methylacyl-CoA Racemase: A new molecular marker for prostate cancer. Cancer Res 2002; 62:2220–2226.

55. Ernst T, Hergenhahn M, Kenzelmann M, Cohen CD, Bonrouhi M, Weninger A, Klaren R, Grone EF, Wiesel M, Gudemann C, Kuster J, Schott W, Staehler G, Kretzler M, Hollstein M, Grone H-J. Decrease and gain of gene expression are equally discriminatory markers for prostate carcinoma: A gene expression analysis on total and microdissected prostate tissue. Am J Pathol 2002;160:2169–2180.

56. LaTulippe E, Satagopan J, Smith A, Scher H, Scardino P, Reuter V, Gerald WL. Comprehensive gene expression analysis of prostate cancer reveals distinct transcriptional programs associated with metastatic disease. Cancer Res 2002;62:4499–4506.

57. Luo JL, Yu YP, Cieply K, Lin F, Deflavia P, Dhir R, Finkelstein S, Michalopoulos G, Becich M. Gene expression analysis of prostate cancers. Mol Carcin 2002;33:25–35.

58. Best CJM, Leiva IM, Chuaqui RF, Gillespie JW, Duray PH, Murgai M, Zhao Y, Simon R, Kang JJ, Green JE, Bostwick DG, Linehan WM, Emmert-Buck MR. Molecular differentiation of high- and moderate-grade human prostate cancer by cDNA microarray analysis. Diagn Mol Pathol 2003;12:63–70.

59. Singh D, Febbo PG, Ross K, Jackson DG, Manola J, Ladd C, Tamayo P, Renshaw AA, D'Amico AV, Richie JP, Lander ES, Loda M, Kantoff PW, Golub TR, Sellers WR. Gene expression correlates of clinical prostate cancer behavior. Cancer Cell 2002;1:203–209.

60. Kristiansen G, Pilarsky C, Wissmann C, Stephan C, Weissbach L, Loy V, Loening S, Dietal M, Rosenthal A. ALCAM/CD166 is upregulated in low-grade prostate cancer and progressively lost in high-grade lesions. Prostate 2003;54:34–43.

61. Velculescu VE, Zhang L, Vogelstein B, Kinzler KW. Serial analysis of gene expression. Science 1995;270:368–369.

62. Xu LL, Su YP, Labiche R, Segawa T, Shanumugam N, McLeod DG, Moul JW, Srivastava S. Quantitative expression profile of androgen-regulated genes in prostate cancer cells and identification of prostate-specific genes. Int J Cancer 2001;92: 322–328.

63. Waghray A, Schober M, Feroze F, Yao Y, Virgin J, Chen YQ. Identification of differentially expressed genes by serial analysis of gene expression in human prostate cancer. Cancer Res 2001; 61:4283–4286.

64. Rubin MA, Zhou M, Dhanasekaran SM, Varambally S, Barrette TR, Sanda MG, Pienta KJ, Ghosh D, Chinnaiyan AM. alpha-Methylacyl coenzyme A racemase as a tissue biomarker for prostate cancer. JAMA 2002;287:1662–1670.