Identification of novel androgen-responsive genes by sequencing of LongSAGE libraries

Tammy L Romanuik, Gang Wang, Robert A Holt, Steven JM Jones, Marco A Marra and Marianne D Sadar*

Address: Genome Sciences Centre, British Columbia Cancer Agency, Vancouver, British Columbia, Canada

Email: Tammy L Romanuik - tromanuik@cw.bcc.ca; Gang Wang - gwang@bcgsc.ca; Robert A Holt - rholt@bcgsc.ca; Steven JM Jones - sjones@bcgsc.ca; Marco A Marra - mmarra@bcgsc.ca; Marianne D Sadar* - msadar@bcgsc.ca

* Corresponding author

Abstract

**Background:** The development and maintenance of the prostate is dependent on androgens and the androgen receptor. The androgen pathway continues to be important in prostate cancer. Here, we evaluated the transcriptome of prostate cancer cells in response to androgen using long serial analysis of gene expression (LongSAGE) libraries.

**Results:** There were 131 tags (87 genes) that displayed statistically significant \((p \leq 0.001)\) differences in expression in response to androgen. Many of the genes identified by LongSAGE (35/87) have not been previously reported to change expression in the direction or sense observed. In regulatory regions of the promoter and/or enhancer regions of some of these genes there are confirmed or potential androgen response elements (AREs). The expression trends of 24 novel genes were validated using quantitative real-time polymerase chain reaction (qRT-PCR). These genes were: ARL6IP5, BLVRB, C19orf48, C1orf122, C6orf66, CAMK2N1, CCNI, DERA, ERRFI1, GLUL, GOLPH3, HM13, HSP90B1, MANEA, NANS, NIPSNAP3A, SLC41A1, SOD1, SVIP, TAOK3, TCP1, TMEM66, USP33, and VTA1. The physiological relevance of these expression trends was evaluated in vivo using the LNCaP Hollow Fibre model. Novel androgen-responsive genes identified here participate in protein synthesis and trafficking, response to oxidative stress, transcription, proliferation, apoptosis, and differentiation.

**Conclusion:** These processes may represent the molecular mechanisms of androgen-dependency of the prostate. Genes that participate in these pathways may be targets for therapies or biomarkers of prostate cancer.

**Background**

Androgens mediate their effect through the androgen receptor (AR) and together they play integral roles in the development and maintenance of the prostate. In the absence of a functional androgen-axis during development, the prostate will fail to form[1]. The size of the prostate increases with the elevation of levels of androgens in males during puberty[2]. Androgens promote proliferation, differentiation, and survival of prostate cells[1]. Men that have used excess androgens in the form of anabolic steroids have a higher incidence of prostate cancer [3-5]. Association of prostate cancer with levels of androgens
has also been reported in rodents[6,7]. Reduction of androgen in humans or dogs before puberty by castration is associated with decreased incidence of prostate cancer[8,9]. Castration of adult males causes apoptosis of prostatic epithelium, involution and reduction of the prostate [10-12]. Thus the prostate gland is an androgen-dependent organ where androgens are the predominant mitogenic stimulus[13]. The dependency of the prostate epithelium on androgens provides the underlying rationale for treating prostate cancer with chemical or surgical castration (androgen-deprivation)[14].

The AR is a ligand-activated transcription factor[15] that regulates transcription of genes that contain androgen response elements (AREs) in the upstream or downstream regulatory regions of the promoter and/or enhancer. Kallikrein 3 (KLK3) is an example of a gene that contains numerous functional AREs that the AR interacts with to increase transcription in response to androgens[16-19]. KLK3, also known as prostate-specific antigen (PSA), is the main tumor marker for prostate cancer and has been used clinically for 15 years[20]. Serum levels of PSA correlate with tumor volume[21]. However, as a screening and monitoring tool for prostate cancer, serum PSA levels are subject to false positives and false negatives[20].

Identification of the genes that change in expression in response to androgen in prostate cells is essential for the understanding of androgen-dependency of the normal prostate and the proliferation, survival, and hormonal progression of prostate cancer. There are several studies that have investigated genes that alter expression in response to a changing androgen-axis using SAGE [22-24]. Here we highlight several key differences in the current experimental design from previous studies: 1) a physiological concentration of metabolically stable androgen (R1881) was employed in vitro; 2) the transcriptome was catalogued using LongSAGE[25] opposed to (short)SAGE[26] because it generates lengthier tags allowing increased confidence in tag-to-gene mapping, and leaves fewer tags unmapped[25]; 3) the transcriptome of human prostate cancer cells was examined instead of murine cells [22]; 4) sequencing depth was increased by approximately 1.5-2 times more tags relative to other studies [23,24] to improve the potential for novel findings; 5) transcript expression was validated using an alternative assay as opposed to protein expression[24], and tens of novel genes were validated as opposed to only two[23]. Thus, we apply LongSAGE for the first time to create transcript libraries of prostate cancer cells maintained in the presence or absence of androgen. These libraries are publicly available at Gene Expression Omnibus. We describe 24 genes never before identified or validated to alter expression in response to androgen treatment. These genes were: ARL6IP5, BLVRB, C19orf48, C1orf122, C6orf66, CAMK2N1, CCNI, DERA, ERRFI1, GLUL, GOLPH3, HM13, HSP90B1, MANEA, NANS, NIPSNAP3A, SLC41A1, SOD1, SVIP, TAOK3, TCP1, TMEM66, USP33, and VTAL. Statistically significant changes in expression of ARL6IP5, CAMK2N1, ERRFI1, HSP90B1, and TAOK3 in response to reduced levels of circulating androgens were measured using in vivo samples.

### Results and discussion

#### Summary of LongSAGE libraries

LongSAGE was employed to obtain quantitative gene expression profiles of human prostate cancer cells treated with or without synthetic androgen R1881. LNCaP human prostate cancer cells were chosen as the model cell line for evaluating androgen signaling because they respond to androgens, express a functional although mutated (T877A) AR, they can be grown in vitro as a monolayer or in vivo as a xenograft or in the Hollow Fiber model [27-29]. LNCaP cells have been used extensively in prostate cancer research. The time of 16 hours for treatment and concentration of R1881 (10 nM) were chosen based upon optimal induction of levels of KLK3 mRNA [30].

LongSAGE libraries were sequenced to a total of 121,760 (R1881) and 103,391 (vehicle) tags (Table 1). The libraries were filtered on several levels to leave only useful tags for analysis. First, bad tags were removed if they contained at least one N-base call in the LongSAGE tag sequence.

| Tag Types            | Total Tags Combined | Unfiltered Total Tags | R1881 | Vehicle |
|----------------------|---------------------|-----------------------|-------|---------|
| No. of Duplicate Ditags | 6,763               | 121,232               | 103,008 |
| % of Duplicate Ditags | 5.579               | 33,385                | 31,764 |
| Average QF of Tags   | 8.964               | 6,763                 | 5,193  |
| No. of Tags QF<95%   | 22,816              | 5,579                 | 5,041  |
| No. of Bad Tags      | 528                 | 8.64                  | 89.67  |
| Average QF           | 8.964               | 8.64                  | 89.67  |
| Total Tags Combined  | 184,329             | 121,760               | 103,391 |
| Tag Types Combined   | 38,574              | 23,830                | 24,594 |
| No. of LDTs Type I   | 219                 | 219                   | 34     |
| No. of LDTs Type II  | 216                 | 216                   | 18     |
| Total Tags Combined  | 183,842             | 183,842               | 183,842 |
| Tag Types Combined   | 38,574              | 38,574                | 38,574 |

* QF, Quality Factor
* s LDTs, Linker-derived Tags

(page number not for citation purposes)
Notably, when bad tags were filtered the percentages of
duplicate ditags in the R1881 and vehicle LongSAGE
libraries were 6% and 5%, respectively. Early SAGE studies
suggest duplicate ditags likely represent polymerase chain
reaction (PCR) artifacts due to the low probability the
same two tags will ligate together to form ditags[26].
However, with LongSAGE library sequencing and highly
expressed transcripts, this probability becomes signifi-
cant[31]. A recent study[32] suggests that discarding
duplicate ditags in LongSAGE analysis may introduce a
bias affecting the fold differences in tag expression
between libraries for all tags observed at a frequency
>(113-224)/100,000. Therefore, we opted to retain dupli-
cate ditags. PHRED software was used to call bases for the
sequencing of the LongSAGE tags[33,34]. PHRED has a
small, but significant error rate in base-calls. To ascertain
which tags potentially contained these erroneous base-
calls, we calculated a tag sequence quality factor (QF) and
probability[35]. The second line of filtering removed
sequences of known sequence were introduced into
SAGE libraries as primers for amplifying ditags prior to
concatenation[26]. These linker sequences were designed
sufficiently unique so they do not map to the human genome. At a low fre-
quence, linkers ligate to themselves creating linker-
derived tags (LDTs). These LDTs do not represent tran-
scripts and are removed from the LongSAGE libraries.
After filtering, there were 97,981 total useful tags repre-
senting 23,828 tag sequences in the R1881 LongSAGE
library, and 85,861 total useful tags representing 24,592
tag sequences in the vehicle LongSAGE library. Due to
redundancy in the expressed sequences, the combined
number of useful tag types in the R1881 and vehicle Long-
SAGE libraries was 38,574. The remainder of the data
analysis in this manuscript was carried out using this fil-
tered data.

**Tag frequency and transcript abundance**

Tag frequency spanned over three orders of magnitude
corresponding to transcript abundance of 5 to 8,746 cop-
ies per cell (based on minimum and maximum observed
tag counts of 1 and 1714; see Table 2 legend)[36]. The distri-
bution of LongSAGE tag frequencies per 100,000 tags
revealed the majority (64 and 67%) of tag types in each
LongSAGE library (R1881 and vehicle, respectively) were
singletons (tags counted only once). This result was con-
sistent with other published SAGE libraries reporting 66%
singletons[37]. Singletons can represent very low abun-
dance transcripts (≤ 5 transcript copies per cell) or PCR/
sequencing errors. Estimates indicate that less than 17% of
LongSAGE tags in a library contain PCR/sequencing errors[38]. Coincidentally, 17% of the total tags in the
R1881 and vehicle LongSAGE libraries roughly equal the
number of singletons in each LongSAGE library (Table 2).
Although initial estimates suggest 6.8-10% of shortSAGE
tags contain PCR/sequencing errors, more recent experi-
mental evidence suggests the actual error rate is much
lower (≤ 2%)[39]. This implies that an error rate of 17% may
also be an overestimate for LongSAGE tags. Tag types
counted 2-4 times per 100,000 tags (10-20 transcript cop-

### Table 2: Characteristics of LongSAGE tag frequency distribution

| Tag Type                  | Total Tags | Tag Types | % of Tags that Map as Transcription Factors | % of Tags Significantly Differentially Expressed |
|---------------------------|------------|-----------|---------------------------------------------|-----------------------------------------------|
| R1881                     | 16,562     | 16,562    | 5.06%                                      | 2.45%                                         |
| Vehicle                   | 16,562     | 16,562    | 5.06%                                      | 2.45%                                         |

| Tag Type                  | Total Tags | Tag Types | % of Tags that Map as Transcription Factors | % of Tags Significantly Differentially Expressed |
|---------------------------|------------|-----------|---------------------------------------------|-----------------------------------------------|
| R1881                     | 16,562     | 16,562    | 5.06%                                      | 2.45%                                         |
| Vehicle                   | 16,562     | 16,562    | 5.06%                                      | 2.45%                                         |

**Tag frequency & Abundance**

- Tag Count per 100,000:
  - ≤1
  - 2-4
  - 5-9
  - 10-99
  - 100-999
  - ≥1,000

- Transcript Copies per Cell:
  - ≤5
  - 10-20
  - 25-45
  - 50-495
  - 500-4,995
  - ≥5,000

- % Transcript Abundance in Cell:
  - ≤0.001
  - 0.002-0.004
  - 0.005-0.009
  - 0.01-0.099
  - 0.1-0.999
  - ≥1

### Notes

**t** Tag count per 100,000 = (observed tag count/total tags in the library) × 100,000

**u** Transcript copies per cell = (observed tag count/total tags in the library) × 500,000

**v** % Transcript abundance in cell = (transcript copies per cell/500,000) × 100%

**w** Calculation based on ~500,000 transcripts in a cell[36]

**a** % of tags that map as transcription factors = (no. of genes with transcription regulation activity/no. of genes with unambiguous sense mappings and GO terms) × 100%

**b** Mapped unambiguously sense to RefSeq and subjected to Gene Ontology (GO) analysis

**c** % of tags from each tag frequency class of R1881 and vehicle LongSAGE libraries were combined

**d** % of tags that map = (no. of tags with sense mappings/combined total tag types) × 100%

**e** Mapped sense (incl. ambiguous) to RefSeq

**f** One tag was mapped sense using Ensembl gene

**g** % of tags significantly differentially expressed = (no. of significantly differentially expressed tag types in class/combined total tag types in class) × 100%

**h** Statistics according to the Audic and Claverie test statistic (p ≤ 0.001)
ies per cell) and 5-9 times per 100,000 tags (25-45 transcript copies per cell) were the second and third most common groups of tag types, respectively. Generally, high frequency tags were less common. The majority of total tags in each LongSAGE library were derived from a few tag types detected between 10-99 times per 100,000 tags (50-495 transcript copies per cell).

Mapping distribution of LongSAGE tags
When mapped tags (v38 Ensembl) were clustered to amalgamate 1-off tags (see Methods, Gene Expression Analysis for a description) and tags that mapped ambiguously were removed, the tag types in the R1881 and vehicle LongSAGE libraries represented 7,484 genes and 7,441 genes, respectively (Table 3). Tag types that mapped ambiguously constituted 13% (R1881 and vehicle), while 36% (R1881) and 35% (vehicle) of tag types did not map to the genome (Table 3). Due to the fact that these tags were clustered, the majority of the tags that did not map to the genome probably represent true unannotated transcripts rather than PCR/sequencing errors. Approximately 28% of tags in each LongSAGE library mapped to the opposite strand of known genes. These LongSAGE tags either represent transcription from previously undescribed coding regions or true antisense transcripts. Each LongSAGE library contained tags representing transcripts from 32% of the genes in the Ensembl gene database. This percentage is indicative of the depth of coverage of the transcriptome achieved with LongSAGE. Alternatively, this percentage indicates that one third of known Ensembl genes were expressed in LNCaP cells under these experimental conditions. This percentage is substantial when considering tag types from the Mouse Atlas Project (8.55 million total LongSAGE tags generated from 72 libraries of mouse development) mapped to 57% of the Ensembl transcript database[35]. Approximately 63% (R1881) and 61% (vehicle) of the genes that mapped to Ensembl's database were associated with more than one tag type to suggest that most gene expression was represented by transcript variants which is consistent with previous observations[35]. When the mapped LongSAGE tags (Reference Sequence, RefSeq; May 18, 2006) were clustered to amalgamate 1-off tags and tags that mapped ambiguously were removed, 53% of tags mapped solely to known exons, 9% solely to known introns (novel transcript variants), and 38% to intergenic regions (novel genes or transcript variants).

The two most abundant tag types in the LongSAGE libraries were shared by both libraries. The first highly abundant LongSAGE tag mapped to human mitochondrial NADH ubiquinone oxidoreductase chain 4. This gene is also highly expressed in other human tissues (i.e., cardiac tissue; SAGE Genie, http://cgap.nih.gov/SAGE). The protein product of this gene transfers electrons from NADH to ubiquinone to generate adenosine triphosphate as metabolic energy. Using the Ensembl database, the second most abundant LongSAGE tag mapped to a non-coding gene of human mitochondria. In contrast to the higher abundance classes, the lower abundance classes were enriched for LongSAGE tags that mapped to genes with functions in regulating transcription (Table 2). This is particularly significant because the percentages of LongSAGE tags that mapped to the genome in the lower abundance class were reduced compared to the higher abundance classes (Table 2). Together this implies that the number of tags that map to genes with a function in transcription may be underestimated, as low abundance tags may be underrepresented.

Differential gene expression
Venn analysis identified that 36% and 38% of tag types were exclusive to the R1881 or vehicle LongSAGE libraries, respectively (Figure 1). The unique expression of tag types indicates differential expression depending upon androgen stimulation. The biological relevance of this differential expression is complicated by the fact that 85% (R1881) and 88% (vehicle) of these exclusive LongSAGE tags were singletons. Consistent with our observation that low abundance tags did not map as readily to the genome, the mutually exclusive tags also did not map as readily as tags shared between both libraries. Only 17% and 15% of tags exclusive to R1881 and vehicle LongSAGE libraries, respectively, mapped unambiguously sense to RefSeq, in contrast to 39% of shared tags. We therefore, concentrated on genes for which the tag abundance allowed for the determination of statistically significant changes in transcript abundance.

Table 3: LongSAGE tag mappings*

| Library | No. of Tag Types that Mapped Unambiguously to (Genes) | No. of Tag Types that Mapped Ambiguously | No. of Tag Types that Did Not Map | Total No. of Tag Types (Clustered) |
|---------|--------------------------------------------------|----------------------------------------|-----------------------------------|-----------------------------------|
| R1881   | 14,587 (7,484)                                    | 3,754                                  | 10,215                            | 28,556                            |
| Vehicle | 13,626 (7,441)                                    | 3,286                                  | 9,066                             | 25,978                            |

x Ensembl gene (v38) was used for mapping
y Clustering amalgamated 1-off tags with likely ‘parent’ tags to improve the mapping capability of LongSAGE tags. Clustering altered the number of tag types without changing the total number of tags in the libraries.
A scatter plot illustrates observed tag counts in LongSAGE libraries relative to the confidence intervals (CIs; 95%, 99%, and 99.9%) of respective p-values (p ≤ 0.05, 0.01, and 0.001) by Audic and Claverie statistics[40] (Figure 2). 891 tags were differentially expressed (p ≤ 0.05) between the two LongSAGE libraries (Figure 2 and Table 4). LongSAGE tags statistically (p ≤ 0.001) differentially represented between the libraries were enriched in the higher abundance classes compared to the lower abundance classes (Table 2). Additionally, 90% of the LongSAGE tags were statistically (p ≤ 0.001) differentially represented between the libraries with ≥ 2-fold differences, compared to only 17% of tags with p-values greater than 0.001 (p > 0.001).

A stringent p-value cutoff (p ≤ 0.001), not corrected for multiple tests, was employed prior to validation of changes in expression of a gene in response to androgen. LongSAGE tags that were differentially expressed, but mapped ambiguously to more than one gene, and/or differenced by less than 2-fold between the treatment groups, were excluded from analysis. Application of these criteria reduced the LongSAGE tags from 131 to 93. These 93 tags represented 87 genes. Analysis of differentially expressed LongSAGE tags revealed that 54 LongSAGE tags that mapped to 52 genes were previously known to change in expression in the direction observed in response to androgen in prostate cancer cells. Of these, the expression of 41 genes increased as expected, including the well-known androgen-regulated gene, \textit{KLK3} (Table 5). The expression of 11 genes decreased in response to androgen and were consistent with previous reports (Table 6). Genes previously not reported to alter expression in response to androgen in prostate cancer cells were represented by 39 LongSAGE tags. These tags represented the expression of 20 genes that were increased, excluding mappings to non-coding and intergenic regions, (Table 7), and expression of 15 genes that was decreased (Table 8) in response to androgen. The 93 tags were represented by 87 genes because one tag did not map to the human genome (Table 7) and two tags mapped to intergenic regions of the human mitochondrial genome (Tables 7 and 8). Three genes were represented twice in the tables (\textit{CAMK2N1}, \textit{PPAP2A}, and \textit{SORD}). One gene, \textit{KRT8}, was categorized in both the known and not previously known categories due to the sense of the mapping (Tables 5 and 8).

Interestingly some antisense tags were identified as differentially expressed in response to androgen. Antisense to \textit{NKX3-1} is of particular note. Transcription of this gene is
Table 5: LongSAGE tags corresponding to genes known to increase expression in response to androgen in LNCaP cells*.

| LongSAGE Tag Sequence | Vehicle | R1881 | Fold \(^c\) \(d\) | Change | RefSeq/Ensembl Access. No. | HGNC Gene Symbol | Description* |
|-----------------------|---------|-------|----------------|--------|--------------------------|-----------------|--------------|
| GTGACAAGTGACAGAGT     | 1       | 19    | 20             |        | NM_007011               | ABHD2           | Abhydrolase domain containing 2, transcript variant 1 |
| ACGTCACCATTITTAAC     | 1       | 24    | 20             |        | NM_004457               | ACSL3           | Acyl-CoA synthetase long-chain family member 3, transcript variant 1 |
| TACTTTATAAGTATTGG     | 14      | 59    | 4.2            |        | NM_006988               | ADAMTS1*        | ADAM metalloproteinase with thrombospondin type 1 motif, 1 |
| TAGCTCTATGGGGGAGG     | 35      | 75    | 2.1            |        | NM_000701               | ATP1A1          | ATPase, Na+/K+ transporting, alpha 1 polypeptide, transcript variant 1 |
| GTTGTGGTTAACTCTG      | 48      | 109   | 2.3            |        | NM_004048               | B2M             | Beta-2-microglobulin |
| ACTTAAGGAACTTATCTT    | 14      | 42    | 3.0            |        | NM_015415               | BRP44           | Brain protein 44 |
| AAAGGAAAAATTTATTATT   | 3       | 27    | 9              |        | NM_018455               | CENPN*          | Centromeric protein N |
| CTGTGATGTGACTCTG      | 5       | 30    | 6              |        | NM_030806               | Clorf21         | Chromosome 1 open reading frame 21 |
| CGATGAGATGTAAGCTG     | 5       | 33    | 7              |        | NM_130898               | CREB3L4*        | cAMP responsive element binding protein 3-like-4 |
| GTGTTATCCTAATCTGA     | 21      | 115   | 5.5            |        | NM_020548               | DEI             | Diazepam binding inhibitor (GAB A receptor modulator, acyl-Coenzyme A binding protein) |
| TCCCCGTGGCTGTGGGG     | 106     | 356   | 3.36           |        | NM_014762               | DHCR24          | 24-dehydrocholesterol reductase |
| GAAATTAGGGAAGCTTTT    | 9       | 34    | 4              |        | NM_015036               | ENDOD1          | Endonuclease domain containing 1 |
| AGATCCTACTTACTGTA     | 16      | 51    | 3.2            |        | NM_004462               | FDF1            | Farnesyl-diphosphate farnesyltransferase 1 |
| GTCCAGCTGAGGGAAGAAG   | 3       | 50    | 20             |        | NM_004117               | FK506*          | FK506 binding protein 5 |
| AACCTAGCCTCTGGGG       | 1       | 24    | 20             |        | NM_002247               | KCNN1           | Potassium large conductance calcium-activated channel, subfamily M, alpha member 1, transcript variant 2 |
| GGATGGGGATGAAGTAA      | 50      | 366   | 7.3            |        | NM_001648               | KLK3*           | Kallikrein 3, (prostate-specific antigen), transcript variant 1 |
| CCTCCAGCTAACAACCA     | 35      | 223   | 6.4            |        | NM_002273               | KRT8            | Keratin 8 |
| TAAATTAGGAAGTTCCTC    | ND*     | 42    | 40             |        | NM_015541               | LRIG1*          | Leucine-rich repeats and immunoglobulin-like domains 1 |
| TCCCTGAGCACCACCTG     | ND*     | 35    | 40             |        | NM_015261               | NCAPD3*         | Non-SMC condensin complex subunit D3 |
| GGAATTCTTCCCTGCCCTA   | 1       | 72    | 70             |        | NM_006096               | NDRG1           | N-myc downstream regulated gene 1 |
| TTTAGGTTAACGGAAAGC    | 19      | 56    | 2.9            |        | NM_014445               | N/A*            | Stress-associated ER protein 1 |
| AGGTGTTTGGCCTATTCC    | 13      | 38    | 2.9            |        | ENSG00000196930         | N/A*            | Similar to Vesicle-associated membrane protein-associated protein A mRNA |
regulated by androgen in a time- and concentration-dependent manner [41] with an ARE confirmed in its

| ATGCAGCCATATGG     | 20 | 208 | 10 | NM_002539 | ODC1 | Ornithine decarboxylase 1 |
|--------------------|----|-----|----|-----------|------|--------------------------|
| GCCAAGGGGCCAGCT    | 17 | 45  | 2.6| NM_002541 | OGDH | Oxoglutarate (alpha-ketoglutarate) dehydrogenase (lipoamide), nuclear gene encoding mitochondrial protein, transcript variant 1 |
|                    |    |     |    |           |      |                          |
| TAAATTTTTACTTTTGT   | 5  | 39  | 8  | NM_017906 | PAK1IP1 | PAK1 interacting protein 1 |
| AC                 |    |     |    |           |      |                          |
| TATGTAATATGCTTTCT   | 27 | 164 | 6.1| NM_003711 | PPAP2A | Phosphatidic acid phosphatase type 2A, transcript variant 1 |
|                    |    |     |    |           |      |                          |
| AAACACCAACAACGGGG   | 5  | 31  | 6  | NM_003711 | PPAP2A | Phosphatidic acid phosphatase type 2A isofoms 1 and 2 |
|                    |    |     |    |           |      |                          |
| GTGTTTACGTGATCCAC   | 1  | 18  | 20 | NM_004578 | RAB4A | RAB4A, member RAS oncogene family |
| CAC                 |    |     |    |           |      |                          |
| TATGTATAATGGACCT    | ND | 16  | 20 | NM_021205 | RHOU* | Ras homolog gene family, member U |
|                    |    |     |    |           |      |                          |
| TTTGAAATGAGGTCTGT   | 14 | 48  | 3.4| NM_002970 | SAT | Spermidine/spermine N1-acetyltransferase Sec61 gamma subunit, transcript variant 1 |
|                    |    |     |    |           |      |                          |
| GCAACAGCAATAGGATT   | 3  | 22  | 7  | NM_014302 | SEC61G | Splicing factor 3B, subunit 5, 10kDa |
|                    |    |     |    |           |      |                          |
| GCGCTGAGGTGAGTGG    | 59 | 126 | 2.1| NM_031287 | SF3B5 | Solute carrier family 45, member 3 |
|                    |    |     |    |           |      |                          |
| GGATTTGAAACATAGA    | ND | 13  | 10 | NM_033102 | SLC4A3 | Sorbitol dehydrogenase |
| AA                 |    |     |    |           |      |                          |
| ACCTTGTGCCCCGATCT   | 47 | 238 | 5.1| NM_003104 | SORD | Sorbitol dehydrogenase |
|                    |    |     |    |           |      |                          |
| AAAATCTGCCACTCAGG   | ND | 12  | 10 | NM_003104 | SORD | Sorbitol dehydrogenase |
|                    |    |     |    |           |      |                          |
| GTGACGGGACGACTCTG    | 3  | 55  | 20 | NM_012391 | SPDEF | SAM pointed domain containing ets transcription factor STEAP family member 4 |
|                    |    |     |    |           |      |                          |
| TTAAGGGATGATGCTTT    | ND | 12  | 10 | NM_024636 | STEAP4 | Transmembrane protein with EGF-like and 2 follistatin-like domains 2 |
|                    |    |     |    |           |      |                          |
| TACTACAGCTATATTG     | 1  | 52  | 3.3| NM_016192 | TMEFF2 | Transmembrane protein, prostate androgen induced RNA, transcript variant 1 |
|                    |    |     |    |           |      |                          |
| TGATGTCGCTGTCGATT    | 1  | 17  | 20 | NM_020182 | TMEPA1 | Transmembrane protease, serine 2 |
|                    |    |     |    |           |      |                          |
| CAAATAAATTATGCGAT    | 5  | 64  | 10 | NM_005656 | TMPRSS2 | Transmembrane protease, serine 2 |
| AT                 |    |     |    |           |      |                          |
| TGAAGGCTTAATATGAT    | 7  | 28  | 4  | NM_005079 | TPDS2 | Tumor protein D52, transcript variant 3 |
| AT                 |    |     |    |           |      |                          |
| TTAAGATTTAGCAACC    | 10 | 36  | 3.6| ENSG00000140416 | TPM1 | Tropomyosin 1 alpha chain |
|                    |    |     |    |           |      |                          |
| TTCTCTACACAATTG     | 6  | 36  | 6  | NM_006022 | TSC22D1 | TSC22 domain family, member 1, transcript variant 1 |
|                    |    |     |    |           |      |                          |

| Table 5: LongSAGE tags corresponding to genes known to increase expression in response to androgen in LNCaP cells* | (Continued) |

*a Statistics according to the Audic and Claverie test statistic (p ≤ 0.001)
*b ND, not detected
*c ND tags were assigned a value of 1 when calculating fold
*d Appropriate significant figures are displayed
*e Ambiguously mapped tags and tags with a fold change less than 2-fold have been excluded from the table
*f N/A = there is no HGNC approved gene symbol for this tag
*g Tag count per 100,000 = (observed tag count/total tags in the library) × 100,000
*h In cases where a tag mapped to >1 transcript variant of the same gene the RefSeq accession number for transcript variant 1 was displayed

* Gene further characterized in this paper
enhancer region [42]. Anti-sense RNA is involved in transcriptional silencing of sense transcript, imprinting control, post-transcriptional down-regulation of sense transcript or even stabilizing/promotion of the expression of the sense transcript [43]. In the case of NKX3-1, anti-sense transcript may be a negative feedback mechanism; however, this remains to be determined.

Validation of changes in gene expression in response to androgen

Quantitative real time-polymerase chain reaction (qRT-PCR) was used to validate changes in gene expression in response to androgen of 39 (13 known; 26 novel) of the 87 total genes identified by LongSAGE. Of the 35 genes previously not reported to change expression in response to androgens in prostate cancer cells, only 26 were quantified by qRT-PCR, because technical limitations and gaps in the transcriptome databases prevented the analysis of 9 genes. That is, specific qRT-PCR primers could not be designed due to repetition in the genome, or because the tag mapped to an unannotated transcript variant.

| LongSAGE Tag Sequence | Vehicle | R1881 | Fold Change | RefSeq/Ensembl Access. No. | HGNC Gene Symbol | Description |
|-----------------------|---------|-------|-------------|--------------------------|-----------------|-------------|
| CAAAAGCTTATTCTTGT     | 29      | 3     | -10         | NM_016613                | C4orf18         | Chromosome 4 open reading frame 18, transcript variant 2 |
| TCACACAGTGCCGTTCG     | 19      | 1     | -20         | NM_020311                | CXCR7*          | Chemokine orphan receptor 1 |
| ACAAACCCCAACCCCAG     | 41      | 7     | -6          | NM_013330                | NME7            | Non-metastatic cells 7, protein expressed in, transcript variant 1, Nucleoside diphosphate kinase |
| AATCTCTCAATTATAGG     | 34      | 9     | -4          | NM_006183                | NTS*            | Neurotensin |
| ATCAAATGGAGGCTCAG     | 15      | ND<sup>b</sup> | -20         | NM_005013                | NUCB2           | Nucleobindin 2 |
| CAAAAATTAGGAAACAC     | 15      | 1     | -20         | NM_002577                | PAK2            | p21 (CDKN1A)-activated kinase 2<sup>k</sup> |
| TTACGTGTGGGAAATAT     | 19      | 2     | -9          | NM_032971                | PCDH11Y         | Protocadherin 11 Y-linked, transcript variant a<sup>k</sup> |
| TGACTTTGGTGCGGTTA     | 12      | ND    | -10         | NM_003629                | PAK3<sup>f</sup> | Phosphoinositide-3-kinase, regulatory subunit 3 (p55, gamma) |
| AGCAAATATGCTAGGG      | 47      | 16    | -2.9        | NM_182948                | PRKACB<sup>k</sup> | Protein kinase, cAMP-dependent, catalytic, beta, transcript variant 1 |
| GACTATTCCATATTAAC     | 27      | 1     | -30         | NM_018412                | ST7<sup>k</sup> | Suppression of tumorigenicity 7, transcript variant A |
| GAGGTTTTAAATGGAGG     | 79      | 9     | -9          | NM_001077                | UGT2B17         | UDP glucuronosyltransferase 2 family, polypeptide B17 |

<sup>a</sup> Statistics according to the Audic and Claverie test statistic (p ≤ 0.001)
<sup>b</sup> ND, not detected
<sup>c</sup> ND tags were assigned a value of 1 when calculating fold change
<sup>d</sup> Appropriate significant figures are displayed
<sup>e</sup> Negative fold change value indicates down-regulation in response to R1881
<sup>f</sup> Tag has a single base pair permutation, insertion, or deletion with respect to gene
<sup>j</sup> Ambiguously mapped tags and tags with a fold change less than 2-fold have been excluded from table
<sup>t</sup> Tag count per 100,000 = (observed tag count/total tags in the library) × 100,000
<sup>φ</sup> If in cases where a tag mapped to >1 transcript variant of the same gene the RefSeq accession number for transcript variant 1 was displayed
<sup>*</sup> Gene further characterized in this paper
Table 7: LongSAGE tags corresponding to genes not previously reported to increase expression in response to androgen in LNCaP cells

| LongSAGE Tag Sequence | Vehicle | R1881 | Fold Change<sup>c</sup> | RefSeq/Ensembl Access. No. | HGNC Gene Symbol | Description<sup>b</sup> |
|-----------------------|---------|-------|-------------------------|---------------------------|------------------|------------------------|
| TCTTTATAGAAAAA AA     | ND<sup>b</sup> | 16    | 20                      | NM_014265                 | ADAM28           | ADAM metallopeptidase domain 28, transcript variant 1<sup>k</sup> |
| AGGAGCAAGGAAG GGG     | 51      | 107   | 2.1                     | NM_000713                 | BLVRB<sup>a</sup> | Biliverdin reductase B (flavin reductase (NADPH)) |
| TTTTGGGGGCTTTTA GC     | 16      | 44    | 2.8                     | NM_198446                 | Clorf122<sup>a</sup> | Chromosome 1 open reading frame 122 |
| GGGCCCAAGCAC TGC     | 22      | 69    | 3.1                     | NM_199249                 | C19orf48<sup>a</sup> | Chromosome 19 open reading frame 48 |
| CCCAGTTGCTGAT CTC     | 24      | 60    | 2.5                     | NM_001003962              | CAPNS1<sup>k</sup> | Calpain, small subunit 1, transcript variant 2 |
| CTTAAGAAATGCA CT      | I       | 23    | 20                      | NM_018948                 | ERFI1<sup>+</sup> | ERBB receptor feedback inhibitor 1 |
| TACAGTATGTTCAAA GT    | 13      | 52    | 4.0                     | NM_002065                 | GLUL<sup>b</sup> | Glutamate-ammonia ligase (glutamine synthetase), transcript variant 1<sup>16</sup> |
| TTAATAGTGGGGGCTTTA TC  | 10      | 39    | 3.9                     | NM_022130                 | GOLPH3<sup>+</sup> | Golgi phosphoprotein 3 (coat protein) |
| GCCAGGCGGGGCA AGC     | ND<sup>b</sup> | 16    | 20                      | NM_178580                 | HMI3<sup>a</sup> | Histocompatibility (minor) 13, transcript variant 2 |
| GAGAAGAAGAAGC AGC     | ND<sup>b</sup> | 14    | 10                      | NM_003299                 | HSP90B1<sup>+</sup> | Heat shock protein 90kDa beta (Grip94), member 1 |
| GGCAAGGGGGGTTCC CCA   | 1       | 20    | 20                      | NM_002273                 | KRT8<sup>a</sup> | Keratin 8<sup>+</sup> |
| ACTCAAAAA AAAAAA AA   | 41      | 81    | 2.0                     | XM_376154                 | N/A<sup>a</sup> | Similar to 40S ribosomal protein S15 (RIG protein), transcript variant 1 |
| GGGTTGGCTTGA AA CCA   | 6       | 30    | 5                       | ENSG00000210151           | N/A<sup>a</sup> | Non-coding predicted mitochondrial gene<sup>+</sup> |
| GAGAGCTCCCGTGAG GTG   | 72      | 122   | 1.7                     | NC_001807<sup>a</sup>    | N/A<sup>a</sup> | Intergenic region of mitochondrial genome N/A |
| TCGGAGCTGATACATCG TTA | 40      | 223   | 5.6                     | No map                    | N/A<sup>a</sup> | |
| GCAAAAAAATCAAG TCT    | 22      | 66    | 3.0                     | NM_018946                 | NANS<sup>a</sup> | N-acetylneuraminic acid phosphate synthase (sialic acid synthase) |
| TCCTTTAGCAAATTC AG     | 2       | 36    | 20                      | NM_006167                 | NKK3<sup>-I</sup> | NK3 transcription factor, locus 1<sup>m</sup> |
| TACCTTTGCGCGTGGC TG    | 6       | 35    | 6                       | NM_173854                 | SLC41A1<sup>a</sup> | Solute carrier family 41, member 1 |
| GAGAGCCTCAAGAAT GGG   | 5       | 26    | 5                       | NM_016281                 | TAOK3<sup>a</sup> | TAO kinase 3 |
| GAAGTTAGAAGAT GCT     | 41      | 106   | 2.6                     | NM_030752                 | TCP1<sup>+</sup> | T-complex protein 1, transcript variant 1 |
| CAGTTCTCTGTGAAA TC     | 40      | 93    | 2.3                     | NM_016127                 | TMEM66<sup>a</sup> | Transmembrane protein 66 |
| ATGGCTTTGTGTGG TT     | ND<sup>b</sup> | 14    | 10                      | NM_201624                 | USP33<sup>a</sup> | Ubiquitin specific protease 33, transcript variant 2 |
DERA, MANEA, NIPSNAP3A, SOD1, SVIP, and VTA1 decreased in response to androgen (Figure 3A). Under the experimental conditions and primers used, we did not measure statistically significant changes in expression of PRNPPIP and CAPNS1. A false discovery rate (FDR) of 29% was expected of the LongSAGE data based on the Audic and Claverie p-value ≤ 0.001. This FDR represents the anticipated percentage of type I errors (i.e., false positives). We observed only 2/26 (8%) false positives, suggesting that the other filter parameters (e.g., ≥ 2-fold difference in expression level) may have increased the chances of validation by qRT-PCR. Moreover, the expression trends for all 13 genes known to change expression in response to androgen in prostate cancer cells correlated between the LongSAGE and qRT-PCR data. ADAMTS1, CENPN, CREB3L4, FKBP5, KLK3, LRG1, NCAPD3, PAK1IP1, and RHOU all increased levels of expression in response to androgen while CXCR7, NTS, PRKACB, and ST7 decreased in response to androgen (Figure 3B).

**Table 7: LongSAGE tags corresponding to genes not previously reported to increase expression in response to androgen in LNCaP cells**

| Tag | Description |
|-----|-------------|
| PRNPPIP | Gene known to change expression in response to androgen, increased expression in PCa cell lines DU145 and PC3 using qRT-PCR |
| CAPNS1 | Gene known to change expression in response to androgen, increased expression in PCa cell lines DU145 and PC3 using qRT-PCR |
| PRKACB | Gene known to change expression in response to androgen, increased expression in PCa cell lines DU145 and PC3 using qRT-PCR |

**Known or potential AREs in the regulatory regions of androgen-regulated genes**

AR directly regulates transcription in response to androgen by binding to AREs in the promoter and/or enhancer regions of target genes. ChIP-chip database mining for suggested AREs combined with a literature search for known AREs revealed some of the genes that alter expression in response to androgen do contain AREs (Table 9). For the 87 genes identified using the cut-off p-value of 0.001 and 2-fold change in response to androgen, there were eight genes with AREs in their promoter, enhancer or intron regions[16,42,45-49]. AREs were detected in the proximity of seven genes by data mining of ChIP-chip studies of ARE on chromosomes 19, 20, 21, 22 [50,51]. Additionally, sequence analysis of the promoters [52] found eight genes from our gene list to contain potential AREs (Table 9). Identification of potential AREs in the regulatory regions of the newly identified genes that alter expression in response to androgen (BLVRB, C19orf48, HM13, SOD1) may be directly regulated by AR.

**Cell-type specificity of gene expression**

To determine if expression of candidate genes was unique to LNCaP cells, we assayed for constitutive levels of expression of 18 known and novel candidate genes in prostate cancer cell lines DU145[53] and PC-3[54] using qRT-PCR (Figure 4). Genes chosen included those that were increased (ADAMTS1, CAPNS1, CENPN, CREB3L4, ERRFI1, FKBP5, HSPI90B1, KLK3, LRG1, NCAPD3, PAK1IP1, and TAOK3) and decreased in expression in response to androgen (ARL6IP5, CAMK2N1, CCNI, CXCR7, PRKACB and ST7). No obvious trends were observed depending on whether expression of the genes increased, or decreased, in response to androgen. All genes tested, except ERRFI1, were expressed at a lower level in PC-3 and DU145 cells relative to LNCaP cells. This suggests that the majority of genes that alter levels of expression in response to androgen were enriched in LNCaP cells relative to PC-3 and DU145 cells. These data are consistent with both DU145 and PC3 cells being androgen-insensitive and lacking a functional AR[53,54].

**In vivo changes in gene expression in response to androgen-deprivation**

The LNCaP Hollow Fibre model combined with qRT-PCR was employed to capture in vivo gene expression representative of physiological levels and castrated levels of androgen (Figure 5). We expected that the genes that had increased levels of expression in vitro in response to androgens, would decrease expression in vivo in response to castration (androgen-deprivation). Conversely, we expected that the genes that had decreased levels of expression in vitro in response to androgens, would increase expression in vivo in response to castration. These in vivo results would be consistent with androgen-responsiveness of the candidate genes. Of the candidate genes examined, 13 of 16 genes showed significant changes in gene expression in response to androgen-deprivation (Figure 5). As anticipated, expression of ARL6IP5, CAMK2N1, CXCR7, and ST7 increased, while CENPN, CREB3L4, ERRFI1, FKBP5, KLK3, LRG1, NCAPD3, PAK1IP1, and TAOK3 decreased...
Table 8: LongSAGE tags corresponding to genes not previously reported to decrease expression in response to androgen in LNCaP cells

| LongSAGE Tag Sequence | Tags/100,000\(^{d,n}\) | Vehicle | R1881 | Fold\(^{-d,j}\) | Change | RefSeq/Ensembl Access. No. | HGNC Gene Symbol | Description\(^{b}\) |
|-----------------------|-------------------------|---------|-------|----------------|--------|---------------------------|-----------------|--------------|
| GTCTAGAATCTGTA CCC    | 29                      |        | 8     | -4             |        | NM_006407                 | ARL6IP5*        | ADP-ribosylation-like factor 6 interacting protein 5 |
| TCAAGAGCCGAAGG AAT    | 12                      | ND\(^{b}\) | -10   | -0.4          |        | NM_014165                 | C6orf66*        | Chromosome 6 open reading frame 66 |
| GTATTGCAAATAATG CC    | 118                     | 24     | -4.9  | -1.2          |        | NM_018584                 | CAMK2NI*        | Calcium/calmodulin-dependent protein kinase II inhibitor 1 |
| AAAAGAAAGCACC TTT     | 30                      | 5      | -6    |              |        | NM_018584                 | CAMK2NI*        | Calcium/calmodulin-dependent protein kinase II inhibitor 1 |
| TTATAACTGAATTTA GT    | 51                      | 11     | -4.6  | -1.1          |        | NM_006835                 | CCNI*           | Cyclin H \(^{h}\) |
| GCAAGGAGAAAGG GAG CAG | 34                      | 7      | -5    |              |        | NP_775809                 | CNBDJ           | N/A \(^{m}\) |
| TGGTACTCATTCAG GC     | 12                      | ND     | -10   | -1.0          |        | NM_015954                 | DERA*           | 2-deoxyribose-5-phosphate aldolase homolog |
| AATCTATGGATTC TT      | 16                      | ND     | -20   |              |        | NM_024641                 | MANEAS*         | Mannosidase, endo-alpha |
| CTAAGACTTACCA GCC     | 19                      | 2      | -10   | -1.0          |        | ENSG00000210082           | N/A           | Non-coding predicted mitochondrial rRNA gene \(^{a}\) |
| CATTGGGTATTTCG TC     | 30                      | 8      | -4    |              |        | NC_001807\(^{p}\)         | N/A            | Intergenic region of mitochondrial genome |
| GTATTTCAGGTCTG TC     | 33                      | 9      | -4    |              |        | NM_015469                 | NIPSAP3A*       | Nipsnap homolog 3A |
| GGTGTTGGTTGGCC CAG    | 23                      | 5      | -5    |              |        | NM_024066                 | PRNPiP*         | Prion protein interacting protein |
| GGTATACCAGCTA AAG     | 122                     | 60     | -2.0  |              |        | NM_002948                 | RPL15           | Ribosomal protein L15 |
| GCACAAGAGATTA AAA     | 58                      | 25     | -2.3  |              |        | NR_002746                 | SNORD47         | Small nuclear RNA, C/D box 47 on chromosome 1 |
| AAAAGACGATGAC TTG     | 77                      | 37     | -2.1  |              |        | NM_000454                 | SOD1*           | Superoxide dismutase I, soluble (amyotrophic lateral sclerosis 1 (adult)) |
| GTTTGGTTATAAATT CT    | 26                      | 3      | -10   |              |        | NM_148893                 | SVIP*           | Hypothetical protein DKFZp313A2432, transcript variant 1 |
| TATTAGAATGAAA AG      | 17                      | 2      | -9    |              |        | NM_016485                 | WA1*            | VPS20-associated 1 homologue |

\(^{a}\) Statistics according to the Audic and Claverie test statistic (p ≤ 0.001)

\(^{b}\) ND, not detected

\(^{c}\) ND tags were assigned a value of 1 when calculating fold change

\(^{d}\) Appropriate significant figures are displayed

\(^{e}\) Gene known to change expression in response to androgens in cells other than prostate

\(^{f}\) Gene known to change expression in response to androgens, but in the opposite direction as reported here

\(^{i}\) Negative fold change value indicates down-regulation by R1881

\(^{k}\) Tag has a single base pair permutation, insertion, or deletion with respect to gene

\(^{l}\) Tag maps to the strand opposite of the gene

\(^{m}\) Ambiguously mapped tags and tags with a fold change less than 2-fold have been excluded from the table

\(^{n}\) N/A = there is no HGNC approved gene symbol for this tag

\(^{p}\) NC_001807 refers to the complete genome of mitochondria in humans

All mitochondrial genes in the RefSeq database are assigned the same accession number by NCBI. q N/A = there is no HGNC approved gene symbol for this tag \(t\) Tag count per 100,000 = (observed tag count/total tags in the library) \(\times\) 100,000

\(^{φ}\) In cases where a tag mapped to >1 transcript variant of the same gene the RefSeq accession number for transcript variant 1 was displayed

\(^{*}\) Gene further characterized in this paper
levels of expression in response to castration. No significant changes in gene expression in vivo were measured for ADAMTS1, HSP90B1, or PRKACB, suggesting that in vivo, other factors may influence their expression. Alternatively, the expression kinetics of each specific gene and half-life of its transcript may vary considerably. The time of harvesting samples and measuring changes in expression of genes in response to androgen-deprivation was at 10 days in vivo compared to 16 hr in vitro in response to addition of androgens (10 nM R1881). Different levels of androgen may also have profound effects on proliferation and differentiation. Physiological levels of androgen in male Nude mice may be considerably lower than the levels used in vitro. Androgen at 10 nM inhibits proliferation of LNCaP cells in vitro while 0.1 nM is optimal for proliferation[55].

**Conclusion**

Androgens are essential for the growth, development and maintenance of the prostate. Here, we created LongSAGE libraries to obtain quantitative gene expression profiles of LNCaP human prostate cancer cells treated with, or without...
out, androgen and revealed the following: 1) 33,385 tag types in the R1881 LongSAGE library and 31,764 tag types in the vehicle LongSAGE library; 2) the majority (64% to 67%) of tag types in each LongSAGE library were singletons which may represent very low abundance transcripts (≤ 5 transcript copies per cell); 3) when mapped tags were clustered and ambiguous mappings were removed, the tag types in the R1881 and vehicle LongSAGE libraries represented 7,484 genes and 7,441 genes, respectively; 4) 53% of tags mapped solely to known exons, 9% solely to known introns (novel transcript variants), and 38% to intergenic regions (novel genes or transcript variants); 5) the most highly abundant LongSAGE tag mapped to human mitochondrial NADH ubiquinone oxidoreductase chain 4 involved in metabolic energy; 6) the lower abundance classes were enriched for genes with functions in regulating transcription; 7) 87 genes were differentially expressed by two-fold (p ≤ 0.001) in response to androgen representing 0.34% of the total tag types (131 differentially expressed tag types/38,574 total tag types); 8) some of these genes have confirmed or potential AREs; 9) novel androgen regulated genes (direct or indirect) identified and validated were

| HGNC Gene Symbol | Access. No. | Expression change by R1881 | Chromosome | Distance of the ARE from TSS (bp) | Reference |
|------------------|------------|----------------------------|------------|----------------------------------|-----------|
| Conventional     |            |                            |            |                                  |           |
| B2M              | NM_004048  | 2.3                        | 15         | -1902                            | [42]      |
| NKX3-1           | NM_006167  | 20                         | 8          | -3013                            | [42]      |
| KLK3             | NM_001649  | 7.3                        | 19         | -170 -4006 -4075 -4115           | [16,48]   |
| FKBPS            | NM_004117  | 20                         | 6          | 65.6 k (Intron 5)                | [47]      |
| NDRG1            | NM_006096  | 70                         | 8          | -984                             | [45]      |
| TMEPAI           | NM_020182  | 20                         | 20         | -2134                            | [45]      |
| TMPRSS2          | NM_005656  | 10                         | 21         | -148                             | [46]      |
| TPDS2            | NM_005079  | 4                          | 8          | -359                             | [49]      |
| Identified by ChIP-Chip |      |                            |            |                                  |           |
| BLVRB            | NM_000713  | 2.1                        | 12         | -56.7 k                          | [51]      |
| C19orf48         | NM_199249  | 3.1                        | 19         | -363 k                           | [51]      |
| CAPN1            | NM_001003962 | 2.5                     | 19         | -165 k                           | [51]      |
| HM13             | NM_178580  | 20                         | 20         | -330 k                           | [51]      |
| ADAMTS1          | NM_006988  | 4.2                        | 21         | 276 k 310 k 481 k                | [50]      |
| TMPRSS2          | NM_005656  | 10                         | 21         | -1063 k                          | [50]      |
| SOD1             | NM_000454  | -2.1                       | 21         | -496 k                           | [50]      |
| Potential based on the sequence |      |                            |            |                                  |           |
| B2M              | NM_004048  | 2.3                        | 15         | -440                             | [52]      |
| NDRG1            | NM_006096  | 70                         | 8          | -1018                            | [52]      |
| NKX3-1           | NM_006167  | 20                         | 8          | -1272                            | [52]      |
| SORD             | NM_003104  | 5.1                        | 15         | -1995                            | [52]      |
| TMEPAI           | NM_020182  | 20                         | 20         | -225                             | [52]      |
| TMPRSS2          | NM_005656  | 10                         | 21         | -771                             | [52]      |
| TPDS2            | NM_005079  | 4                          | 8          | -609                             | [52]      |
| TSC22D1          | NM_006022  | 6                          | 13         | -1711                            | [52]      |

*TSS: Transcription starting site

Table 9: Genes with confirmed or potential AREs that change expression in response to androgen

- **B2M**: NM_004048
- **NKX3-1**: NM_006167
- **KLK3**: NM_001649
- **FKBPS**: NM_004117
- **NDRG1**: NM_006096
- **TMEPAI**: NM_020182
- **TMPRSS2**: NM_005656
- **TPDS2**: NM_005079
- **B2M**: NM_004048
- **NDRG1**: NM_006096
- **NKX3-1**: NM_006167
- **SORD**: NM_003104
- **TMEPAI**: NM_020182
- **TMPRSS2**: NM_005656
- **TPDS2**: NM_005079
- **B2M**: NM_004048
- **NDRG1**: NM_006096
- **NKX3-1**: NM_006167
- **SORD**: NM_003104
- **TMEPAI**: NM_020182
- **TMPRSS2**: NM_005656
- **TPDS2**: NM_005079
- **B2M**: NM_004048
- **NDRG1**: NM_006096
- **NKX3-1**: NM_006167
- **SORD**: NM_003104
- **TMEPAI**: NM_020182
- **TMPRSS2**: NM_005656
- **TPDS2**: NM_005079

*TSS: Transcription starting site
Figure 4
Differential expression of candidate genes in LNCaP, DU145, and PC-3 cells. Levels of transcripts in LNCaP, DU145, and PC-3 cells were analyzed by qRT-PCR. Error bars represent ± standard deviation (SD) for biological triplicates. [*] Asterisks indicate the significant differential gene expression in each cell line compared to LNCaP cells according to the Two-Sample Student’s T-test (p ≤ 0.05) for equal (unpaired) or unequal variance as determined appropriate with the F-test.
**Figure 5**

Androgen regulation of genes in the in vivo Hollow Fibre model of prostate cancer. Levels of transcripts in LNCaP cells from the Hollow Fibre model were analyzed by qRT-PCR. Cx, castrated mice, 10 days post castration, n = 12; Pre-Cx, pre-castration, day 0 of castration, n = 15. Exception: LRIG1 gene expression in Cx samples was represented by 11 mice. Fold-change was calculated by normalizing the mean normalized expression (MNE) of transcripts in the Pre-Cx sample to the castrate sample. In doing this, the Cx sample fold-change became one and standard deviation (SD) zero. Error bars represent ± SD. [*] Asterisks indicate the significant differential gene expression with respect to Cx according to the Two-Sample Student’s T-test (p ≤ 0.05) for unequal variance.

| Gene       | Pre-Cx | Cx  | Fold Change | Pre-Cx | Cx  | Fold Change | Pre-Cx | Cx  | Fold Change | Pre-Cx | Cx  | Fold Change | Pre-Cx | Cx  | Fold Change |
|------------|--------|-----|-------------|--------|-----|-------------|--------|-----|-------------|--------|-----|-------------|--------|-----|-------------|
| ADAMTS1    |        |     |             |        |     |             |        |     |             |        |     |             |        |     |             |
| ARL6IP5    |        |     |             |        |     |             |        |     |             |        |     |             |        |     |             |
| CAMK2N1    |        |     |             |        |     |             |        |     |             |        |     |             |        |     |             |
| CENPN      |        |     |             |        |     |             |        |     |             |        |     |             |        |     |             |
| CREB3L4    |        |     |             |        |     |             |        |     |             |        |     |             |        |     |             |
| CXCR7      |        |     |             |        |     |             |        |     |             |        |     |             |        |     |             |
| ERRFI1     |        |     |             |        |     |             |        |     |             |        |     |             |        |     |             |
| FKBP5      |        |     |             |        |     |             |        |     |             |        |     |             |        |     |             |
| HSP90B1    |        |     |             |        |     |             |        |     |             |        |     |             |        |     |             |
| KLK3       |        |     |             |        |     |             |        |     |             |        |     |             |        |     |             |
| LRIG1      |        |     |             |        |     |             |        |     |             |        |     |             |        |     |             |
| NCAPD3     |        |     |             |        |     |             |        |     |             |        |     |             |        |     |             |
| PAK1IP1    |        |     |             |        |     |             |        |     |             |        |     |             |        |     |             |
| PRKACB     |        |     |             |        |     |             |        |     |             |        |     |             |        |     |             |
| ST7        |        |     |             |        |     |             |        |     |             |        |     |             |        |     |             |
| TAOK3      |        |     |             |        |     |             |        |     |             |        |     |             |        |     |             |

CXCR7, FKBP5, HSP90B1, KLK3, LRIG1, NCAPD3, PAK1IP1, PRKACB, ST7, and TAOK3 was increased in LNCaP cells compared to prostate cancer cells lacking a functional AR; and 10) significant differences in levels of expression of ARL6IP5, CAMK2N1, CENPN, CREB3L4, CXCR7, ERRFI1, FKBP5, KLK3, LRIG1, NCAPD3, PAK1IP1, ST7, and TAOK3 were measured in vivo in response to androgen-deprivation. The products of these genes are involved in amino acid and protein synthesis, cofactor transport, protein trafficking, response to oxidative stress, as well as signaling pathways that regulate gene expression, proliferation, apoptosis, and differentiation.

These genes are potentially critical for the function and maintenance of the prostate and represent targets for clinical intervention.

**Methods**

**Cell culture**

LNCaP human prostate cancer cells (American Type Culture Collection, Bethesda, MD, USA) were maintained in RPMI-1640 media (Stem Cell Technologies, Vancouver, BC, Canada) supplemented with 10% v/v fetal bovine serum (FBS; HyClone, Logan, UT, USA), 100 units/mL penicillin and 100 units/mL streptomycin (antibiotics;
Invitrogen, Burlington, ON, Canada). DU145 and PC-3 human prostate cancer cells were maintained in DMEM (Stem Cell Technologies) supplemented with 10% v/v FBS and 5% v/v FBS, respectively with antibiotics. All cells were maintained at 37°C with 5% CO2.

Long serial analysis of gene expression
RNA sample generation
1 × 10^6 LNCaP cells were seeded in 10 cm-diameter dishes. The next day, cells were serum-starved (0% serum) for 48 hours and then treated for 16 hours with 10 nM synthetic androgen R1881 (also known as methyltrienolone; PerkinElmer, Woodbridge, ON, Canada), or solvent (vehicle) control, ethanol (final concentration 2.85 × 10^-4%). Total RNA was extracted using TRIZOL Reagent (Invitrogen) following the manufacturer’s instructions. RNA quality and quantity were assessed using the Agilent 2100 Bioanalyzer (Agilent Technologies, Mississauga, ON, Canada) and RNA 6000 Nano LabChip kit (Caliper Technologies, Hopkinton, MA, USA).

LongSAGE library production
LongSAGE[25] libraries were constructed with 5 μg of total RNA using the Invitrogen I-SAGE Long kit and protocol with alterations as previously published[35]. Briefly, double-stranded cDNA was synthesized from total RNA and digested with Nla III. The sample was split in half and linkers type I and II were added and ligated to Nla III overhangs. An Mme I digestion resulted in 17-21 base-pair (bp) LongSAGE tags. The tags with unique linkers were combined and ligated together to form ditags. Ditags (131 bp) were amplified with primers designed to recognize sequences within linkers type I and II using PCR. This scale-up PCR was performed in 48 wells of a 96 well plate (50 μL/well) using a 1/20th dilution of template cDNA and 25 and 27 cycles of PCR (R1881 and vehicle LongSAGE library, respectively). Following an Nla III digestion to remove the linkers, the 36 bp ditags were concatenated. Concatemers sized 1300-1700 bp were digested with Nla III (1 minute) to increase the efficiency of cloning into pZErO-1 vectors. Cloned concatemers were transformed into One Shot TOP10 Electrocompetent Escherichia coli and colonies were picked with the Q-Pix robot (Genetix) and cultured in 2× Yeast-Tryptone media with 50 μg/mL zeocin and 7.5% (v/v) glycerol.

Sequencing
Glycerol stocks of transformed bacteria were used to inoculate larger cultures for alkaline lysis plasmid preparation[56]. Plasmid preparations were separated by agarose gel electrophoresis and visualized by ultraviolet light and sybr green. 1/24th BigDye v3.1 terminator cycle sequencing reactions were performed with tetrat thermal cyclers (BioRad, Waltham, MA, USA) and visualized with capillary DNA sequencers, models 3700 and 3730 xl (Applied Biosystems, Foster City, CA, USA). Each library was sequenced to a depth of ~100,000 LongSAGE tags. Flanking vector sequences were removed and the LongSAGE tags were extracted from each sequence read. On average, 34 and 38 LongSAGE tags were sequenced in each read (R1881 and vehicle libraries, respectively). Sequence data were filtered for non-recombintant clones.

Gene expression analysis
LongSAGE expression data was analyzed with DiscoverSpace 3.2.4 and 4.01 software http://www.bcgsc.ca/bioinfo/software/discoveryspace/. Duplicate ditags (identical copies of a ditag) and singletons (tags counted only once) were retained for analysis. Sequence data were filtered for bad tags (tags with one N-base call) and linker-derived tags (artifact tags). Only LongSAGE tags with a sequence quality factor (QF) greater than 95% were included in analysis[35]. Where indicated, a clustering algorithm was used to amalgamate 1-off tags (tags one bp incorrect from a complete map to a transcript) with likely ‘parent' tags to improve the mapping capability of LongSAGE tags by apparently reducing PCR/sequencing errors[35]. This clustering algorithm altered the number of tag types (i.e., species) without changing the total number of tags. In instances where clustering was used, the 95% QF cutoff was not. To filter data for candidate transcript validation, a p-value cutoff (p ≤ 0.001) was employed according to the Audic and Claverie test statistic[40]. The Audic and Claverie statistical method was used to identify differentially expressed tags between LongSAGE libraries because the method takes into account the sizes of the libraries and tag counts. LongSAGE tags that mapped ambiguously to more than one gene, and tags that differed by less than 2-fold were excluded from the candidate list. LongSAGE tags were mapped to reference sequence (RefSeq; May 30th, 2005) and Ensembl Gene (v31.35d), unless otherwise stated.

Quantitative real-time polymerase chain reaction
qRT-PCR was performed on TRIZOL-extracted RNA from LNCaP (serum-starved ± R1881 or the exception in Figure 4 in 10% serum), DU145 (10% serum) and PC-3 (5% serum) cells maintained in vitro, and LNCaP cells maintained in the in vivo Hollow Fibre model[29] (see below). Contaminating genomic DNA was removed from in vitro RNA samples using DNA-free or TURBO DNA-free (Ambion, Austin, TX, USA). Input RNA (1 μg) was reverse transcribed with SuperScript III First Strand Synthesis kit (Invitrogen). A 10 μL qRT-PCR reaction included 1 μL of template cDNA (0.1 μL for limited LNCaP Hollow Fibre samples), 1× Platinum SYBR Green qPCR SuperMix-UDG with ROX (Invitrogen) and 0.3 μM each of forward and reverse intron-spanning primers that produce products between 85-115 bp in size (see Additional file 1 for primer sequences). qRT-PCR reactions were cycled as follows in a
were contaminated with mouse cells, as indicated by an

cells harvested from the fibres. Compromised fibres that

four fibres). Total RNA was isolated immediately from

removed at day zero (Pre-Cx; four fibres) and day 10 (Cx;

laboratories, Abbott Park, IL, USA). Bundles of fibres were

determined by enzymatic immunoassay kit (Abbott Labo-

trum the response to castration. Serum KLK3 levels were

tail vein each week to measure serum KLK3 levels to mon-

trum; VTA1: vps20-associated 1.

HM13: Histocompatibility (minor) 13; HSP90B1: heat

mate-ammonia ligase; GOLPH3: golgi phosphoprotein 3;

KLK3: kallikrein 3

Abbreviations
AR: androgen receptor; ARES: androgen response ele-

ments; ARL6IP5: ADP-ribosylation like factor-6 interact-

ing protein 5; CAMK2N1: calcium/calmodulin-depen-

dent protein kinase II inhibitor 1; CI: confidence

interval; Cx: castration; ERRF11: ERBB receptor feedback

inhibitor; FBS: fetal bovine serum; FDR: false discovery

rate; GAPDH: glyceraldehyde-3-phosphate; GLUL: gluta-

mate-ammonia ligase; GOLPH3: golgi phosphoprotein 3;

H13: Histocompatibility (minor) 13; HSP90B1: heat

shock protein 90 kDa beta member 1; KLK3: kallikrein 3

= PSA; LDT: linker-derived tag; LongSAGE: long serial

analysis of gene expression; MANEA: mannosidase: endo

alpha; MNE: mean normalized expression; NANS: n-

acetylneuraminic acid synthase; NIPSNAP3A: nipsnap

homologue 3A; PCR: polymerase-chain reaction; PSA:

prostate-specific antigen = KLK3; QF: quality factor; qRT-

PCR: quantitative real-time polymerase chain reaction;

R1881: methyltrienolone: synthetic androgen; RefSeq:

reference sequence; SAGE: serial analysis of gene expres-

sion; SD: standard deviation; shortSAGE: short serial anal-

ysis of gene expression; SLC41A1: solute carrier family 41:

member 1; SOD1: superoxide dismutase 1; SWP: small

VCP/p97-interacting protein; TAOK3: tao kinase 3; TCP1:

T-complex 1; VIA1: vps20-associated 1.

Authors' contributions
TLR conducted the experiments, analyzed the data and

wrote the manuscript. GW generated the total RNA, ana-

lyzed the data, and helped to draft the manuscript. MAM

provided support for the SAGE library construction with

sequencing by RAH. SIM aided in the analysis of data. MDS

conceived the study, designed the experiments, and

coordination and wrote the manuscript. All authors read

and approved the final manuscript.

Author's Information
Tammy L. Romanuik, PhD: Graduate Student, Genome

Sciences Centre, BC Cancer Agency

Current address: Medical Genetics, UBC

Gang Wang, PhD: Postdoctoral fellow, Genome Sciences

Centre, BC Cancer Agency

Robert A. Holt, PhD: Head of Sequencing, Genome Sciences

Centre, BC Cancer Agency

Steven J. M. Jones, PhD: Associate Director/Head of Bioin-

formatics, Genome Sciences Centre, BC Cancer Agency

Marco A. Marra, PhD: Director, Genome Sciences Centre
Additional material

Additional file 1
Primer sequences and amplification product sizes for candidate transcripts. The data provided represent the primer sequences used in quantitative real-time polymerase chain reaction to validate changes in gene expression in response to androgen.

Click here for file
[http://www.biomedcentral.com/content supplemen tary/1471-2164-10-476-S1.PDF](http://www.biomedcentral.com/content/supplementary/1471-2164-10-476-S1.PDF)

Acknowledgements

The authors would like to thank Jean Wang, Nasrin (Rina) Mawji, and Angelique Schnerr for their excellent technical assistance. This work is supported by funding from NIH, Grant CA105304 (M.D.S.). M.A.M. and S.J.M. are Senior Scholars of the Michael Smith Foundation for Health Research. M.D.S. and M.A.M. are Terry Fox Young Investigators. S.J.M. is Senior Early Career Scholar of the Peter Wall Institute for Advanced Studies.

LongSAGE libraries are available at Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/); series accession number GSE18401; R1881 sample accession number GSM458900; Vehicle sample accession number GSM458901.

References

1. Cunha GR, Ricke W, Thomson A, Marker PC, Risbridger G, Hayward SW, Wang YZ, Donjacour AA, Kurita T: Hormonal, cellular, and molecular regulation of normal and neoplastic prostate development. J Steroid Biochem Mol Biol 2004, 92(4):221-236.
2. Yong EL, Lim J, Qi W, Ong V, Mifsud A: Molecular basis of androgen receptor diseases. Ann Med 2000, 32(1):15-22.
3. Guinan PD, Sadouh W, Alsheik H, Ablin RJ, Alrenga D, Bush IM: Impotence therapy and cancer of the prostate. Ann Surg 1976, 113(5):599-600.
4. Jackson JA, Waxman J, Spiekerman AM: Prostatic complications of testosterone replacement therapy. Arch Intern Med 1989, 149(10):2365-2366.
5. Roberts JT, Essenhig DM: Adenocarcinoma of prostate in 40-year-old body-builder. Lancet 1986, 2(8509):742.
6. Noble RL: The development of prostatic adenocarcinoma in Nb rats following prolonged sex hormone administration. Cancer Res 1977, 37(6):1929-1933.
7. Noble RL: Sex steroids as a cause of adenocarcinoma of the dorsal prostate in Nb rats, and their influence on the growth of transplants. Oncology 1977, 34(3):138-141.
8. Wilding G: The importance of steroid hormones in prostate cancer. Cancer Surv 1992, 14:113-130.
9. Wilson JD, Roehrborn C: Long-term consequences of castration in men: lessons from the Skopky and the eunuchs of the Chinese and Ottoman courts. J Clin Endocrinol Metab 1999, 84(12):4324-4331.
10. Bruckheimer EM, Kyriakou N: Apoptosis in prostate carcinogenesis. A growth regulator and a therapeutic target. Cell Tissue Res 2000, 301(1):153-162.
11. Isaacs JT: Antagonistic effect of androgen on prostatic cell death. Prostate 1984, 5(5):545-557.
12. Wu CP, Gu FL: The prostate in eunuchs. Prog Clin Biol Res 1991, 370:249-257.
13. Isaacs JT, Scott WW, Coffey DS: New biochemical methods to determine androgen sensitivity of prostatic cancer: the relative enzymatic index (REI). Prog Clin Biol Res 1979, 33:133-144.
14. Huggins C, Hodges C: Studies on prostatic cancer: The effect of castration, of estrogen and of androgen injection on serum phosphatases in metastatic carcinoma of the prostate. Cancer Res 1941(1):293-297.
15. Yamamoto KR: Steroid receptor regulated transcription of specific genes and gene networks. Annu Rev Genet 1985, 19:209-252.
16. Cleutjens KB, Korpút HA van der, van Eekelen CC, van Rooij HC, Faber PW, Trapman J: An androgen response element in a far upstream enhancer region is essential for high, androgen-regulated activity of the prostate-specific antigen promoter. Molecular endocrinology (Baltimore, Md) 1997, 11(2):148-161.
17. Song Y, Myers M, Brown M: Formation of the androgen receptor transcription complex. Molecular cell 2002, 9(3):601-610.
18. Wolf DA, Schulz P, Fidler F: Transcriptional regulation of prostate kallikrein-like genes by androgen. Molecular endocrinology (Baltimore, Md) 1992, 6(5):753-762.
19. Schwartz ER, Henderson GA, Kmetec LA, Miller JD, Lamparski HG, Henderson DR: Prostate-specific antigen expression is regulated by an upstream enhancer. J Biol Chem 1996, 271(12):7043-7051.
20. Small EJ: Prostate cancer: who to screen, and what the results mean. Genetics 1993, 138(2-3):599-600.
21. Grossklaus D, Shappell SB, Gatum S, Smith JA Jr, Cookson MS: Ratio of free-to-total prostate specific antigen correlates with tumor volume in patients with increased prostate specific antigen. J Urol 2001, 165(2):455-458.
22. Chi M, Mayumi Y, Andre B, Pascale B, Lin G, Yasukazu T, Fernand L, Jonny SA: Prostate-specific genes and their regulation by dihydrotestosterone. Prostate 2008, 68(3):241-254.
23. Xu LL, Su YP, Labiche T, Segawa T, Shanmugam 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(3):322-328.
24. Waghary A, Feroze F, Schober MS, Yao F, Wood C, Purvus E, Krause M, Hanash S, Chen YQ: Identification of androgen-regulated genes in the prostate cancer cell line LNCaP by serial analysis of gene expression and proteomic analysis. Proteomics 2001, 1(10):1327-1338.
25. Saha S, Sparks AB, Rago C, Akmaev V, Wang CJ, Vogelstein B, Kinzler KW, Velculescu VE: Using the transcriptome to annotate the genome. Nat Biotechnol 2002, 20(5):508-512.
26. Velculescu VE, Zhang L, Vogelstein B, Kinzler KW: Serial analysis of gene expression. Science 1995, 270(5235):484-487.
27. Gleaves M, Hsieh JT, Gao CA, von Eschenbach AC, Chung LW: Acceleration of human prostate cancer growth in vivo by factors produced by prostate and bone fibroblasts. Cancer Res 1991, 51(14):3753-3761.
28. Horoszewicz JS, Leong SS, Chu TM, Wajsman ZL, Friedman M, Papasimakis P, Dhalla N, Prabhu AL, Ma K, et al.: Large-scale production of SAGE libraries from microdissected tissues, flow-sorted cells, and cell lines. Genome Res 2007, 17(1):108-116.
29. Emmersen J, Heidenblut AM, Hough AL, Hahn SA, Welinder KG, Nielsen KL: Discarding duplicate SAGE tags in LongSAGE analysis may introduce significant error. BMC Bioinformatics 2007, 892.
30. Ewing B, Green P: Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res 1998, 8(3):186-194.
31. Ewing B, Hillier L, Wendl MC, Green P: Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res 1998, 8(3):173-185.
32. Siddiqui AS, Khattra J, Delaney AD, Zhao Y, Astell C, Asano J, Babakaff R, Barber S, Beland J, Bohacek S, et al.: A mouse atlas of gene expression: large-scale digital gene-expression profiles
from precisely defined developing C57BL/6 mouse tissues and cells. Proc Natl Acad Sci USA 2005, 102(51):18485-18490.
36. Hastie ND, Bishop JO: The expression of three abundance classes of messenger RNA in mouse tissues. Cell 1976, 9(4):PT 2):761-774.
37. Margulies EH, Kardia SL, Innis JW: A comparative molecular analysis of developing mouse forelimbs and hindlimbs using serial analysis of gene expression (SAGE). Genome Res 2001, 11(10):1686-1698.
38. Akmaev VR, Wang CJ: Correction of sequence-based artifacts in serial analysis of gene expression. Bioinformatics 2004, 20(8):1254-1263.
39. Chen J, Sun M, Lee S, Zhou G, Rowley JD, Wang SM: Identifying novel transcripts and novel genes in the human genome by using novel SAGE tags. Proc Natl Acad Sci USA 2002, 99(19):12257-12262.
40. Audic S, Claverie JM: The significance of digital gene expression profiles. Genome Res 1997, 7(10):986-995.
41. Prescott JL, Blok L, Tindall DJ: Isolation and androgen regulation of the human homeobox cDNA, NKX3.1. Prostate 1998, 35(1):71-80.
42. Yoon HG, Wong J: The corepressors silencing mediator of retinoid and thyroid hormone receptor and nuclear receptor corepressor are involved in agonist- and antagonist-regulated transcription by androgen receptor. Molecular endocrinology (Baltimore, Md) 2006, 20(5):1048-1060.
43. Beiter T, Reich E, Williams RW, Simon P: Antisense transcription: a critical look in both directions. Cell Mol Life Sci 2009, 66(1):94-112.
44. Storey JD: A direct approach to false discovery rates. JR Statist Soc B 2002, 64:479-498.
45. Kim KT, Dobi A, Shaheduzzaman S, Gao CL, Masuda K, Li H, Druckier A, Gu Y, Srikantan V, Rhim JS, et al.: Characterization of the androgen receptor in a benign prostate tissue-derived human prostate epithelial cell line: RC-165N/human telomerase reverse transcriptase. Prostate cancer and prostatic diseases 2007, 10(1):30-38.
46. Lin B, Ferguson C, White JT, 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(17):4180-4184.
47. Pegg EA, Chang LW, Stormo GD, Milbrandt J: Direct, androgen receptor-mediated regulation of the FKBP5 gene via a distal enhancer element. Endocrinology 2006, 147(1):590-598.
48. Riehm PH, Vlieistra RJ, Korpot JA van der, Brinkmann AO, Trapman J: The promoter of the prostate-specific antigen gene contains a functional androgen responsive element. Molecular endocrinology (Baltimore, Md) 1991, 5(12):1921-1930.
49. Wang R, Xu J, Saramaki O, Visakorpi T, Sutherland WM, Zhou J, Sen B, Lim SD, Mabjeesh N, Amin M, et al.: ProLZ, a novel prostate-specific and androgen-responsive gene of the TPD52 family, amplified in chromosome 8q21.1 and overexpressed in specific and androgen-responsive gene of the TPD52 family, Prostate 1998, 71-80.
50. Wang Q, Li W, Liu XS, Carroll JS, Janne OA, Chinnaiyan AM, Pienta KJ, Brown M: A hierarchical network of transcription factors governs androgen receptor-dependent prostate cancer growth. Molecular cell 2007, 27(3):380-392.
51. Jia L, Berman BP, Jariwala U, Yan X, Cogan JP, Walters A, Chen T, Buchanan G, Frenkel B, Coetsee GA: Genomic androgen receptor-occupied regions with different functions, defined by histone acetylation, coregulators and transcriptional capacity. Proc Natl Acad Sci USA 2008, 105(11):3645.
52. 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. Proc Natl Acad Sci USA 2002, 99(18):11890-11895.
53. Stone KR, Mickey DD, Wunderli H, Mickey GH, Paulson DF: Isolation of a human prostate carcinoma cell line (DU 145). Int J Cancer 1978, 21(3):274-281.
54. Kaighn ME, Narayan KS, Ohnuki Y, Lechner JF, Jones LW: Establish-ment and characterization of a human prostatic carcinoma cell line (PC-3). Invest Urol 1979, 17(1):16-23.
55. Yang GS, Scott JM, Smalldus D, Barber SA, Balasundaram M, Marra M, Holt RA: High-throughput sequencing: a failure mode analysis. BMC Genomics 2005, 6(1):2.
56. Robertson N, Oveis-Fordore M, Zuyderduyn SD, Varhol R, Fjell C, Marra M, Jones S, Siddiqui A: DiscoverySpace: an interactive data analysis application. Genome Biol 2007, 8(1):R6.