

REPORT

A network biology approach to prostate cancer

Ayla Ergün¹, Carolyn A Lawrence¹, Michael A Kohanski^{1,2}, Timothy A Brennan¹ and James J Collins^{1,*}

¹ Department of Biomedical Engineering and Center for BioDynamics, Boston University, Boston, MA, USA and ² Boston University School of Medicine, Boston, MA, USA

* Corresponding author. Department of Biomedical Engineering and Center for BioDynamics, Boston University, 44 Cummington Street, Boston, MA 02215, USA. Tel.: +1 617 353 0390; Fax: +1 617 353 5462; E-mail: jcollins@bu.edu

Received 1.8.06; accepted 22.12.06

There is a need to identify genetic mediators of solid-tumor cancers, such as prostate cancer, where invasion and distant metastases determine the clinical outcome of the disease. Whole-genome expression profiling offers promise in this regard, but can be complicated by the challenge of identifying the genes affected by a condition from the hundreds to thousands of genes that exhibit changes in expression. Here, we show that reverse-engineered gene networks can be combined with expression profiles to compute the likelihood that genes and associated pathways are mediators of a disease. We apply our method to non-recurrent primary and metastatic prostate cancer data, and identify the androgen receptor gene (AR) among the top genetic mediators and the AR pathway as a highly enriched pathway for metastatic prostate cancer. These results were not obtained on the basis of expression change alone. We further demonstrate that the AR gene, in the context of the network, can be used as a marker to detect the aggressiveness of primary prostate cancers. This work shows that a network biology approach can be used advantageously to identify the genetic mediators and mediating pathways associated with a disease.

Molecular Systems Biology 13 February 2007; doi:10.1038/msb4100125

Subject Categories: metabolic and regulatory networks; molecular biology of disease

Keywords: disease mediators; prostate cancer; reverse-engineered gene networks; systems biology

Introduction

Since the introduction of microarrays, there has been considerable interest in using whole-genome expression profiling to gain insight into cancer and to identify key genetic mediators. Although there has been some success in this regard, many efforts have been complicated by the fact that it is difficult with expression data to identify the genes affected by a condition from the hundreds to thousands of genes that exhibit changes in expression. In this work, we show that a network biology approach can be used to take on this challenge. Specifically, we show that reverse-engineered gene networks can be combined with expression profiles to identify the genetic mediators and mediating pathways associated with prostate cancer.

We used an approach called mode-of-action by network identification (MNI), which has previously been validated as a means to identify the targets and associated pathways of compounds (di Bernardo *et al.*, 2005). The MNI algorithm operates in two phases (Figure 1). In phase one, a network model of regulatory interactions is reverse engineered with a diverse training set of whole-genome expression profiles. In phase two, the network is used as a filter to determine the genes affected by a condition of interest, for example, a disease (Figure 1). The highest ranked mediator genes, ranked by a

Z-statistic, are those whose expression is most inconsistent with the model, and this inconsistency is attributed to the external influence of the condition on those genes. Genes implicated in the advancement as well as suppression of a disease are equally likely to be identified as significant genetic mediators by the MNI algorithm (see Supplementary information). The MNI algorithm requires that the training expression profiles influence a diversity of cell functions. As a training set, we used a total of 1144 microarray expression profiles based on 13 projects spanning seven different cancer types: adrenal, brain, breast, leukemia, lung, prostate and thyroid (Materials and methods). As test conditions, we used expression profiles of 14 non-recurrent primary and nine distant metastatic prostate cancer samples (LaTulippe *et al.*, 2002). Each of these samples was queried against the reconstructed network, and the resulting potential genetic mediators in each case were ranked according to the Z-score statistic. A characteristic list of 100 genes for the non-recurrent primary and metastatic prostate cancer groups, respectively, was obtained by averaging the Z-scores across all samples.

Normal prostate and early-stage prostate cancers depend on androgens for growth and survival. As the cancer advances and metastasizes, it becomes dominated by cells that proliferate and survive independent of androgens; this effect is provoked by androgen ablation therapy (Taplin *et al.*, 1995;

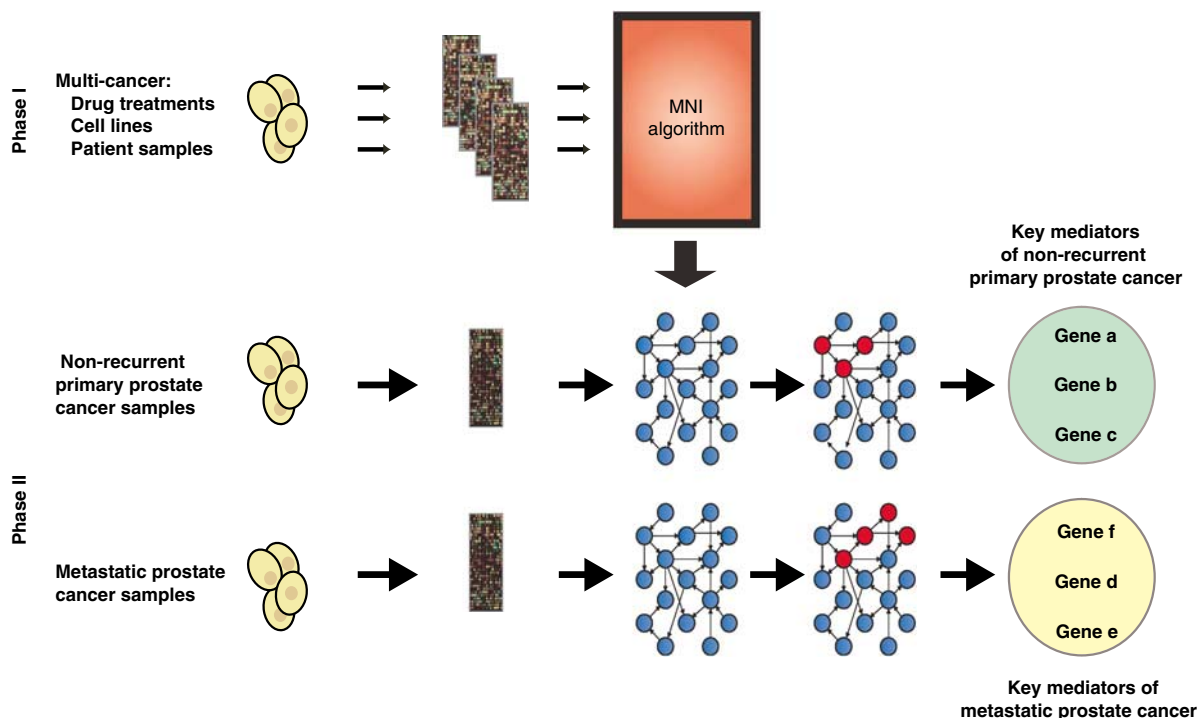


Figure 1 A schematic diagram of the MNI method as applied to identify genetic mediators for non-recurrent primary and metastatic prostate cancer. In phase 1, microarray data obtained from a variety of cancer cell lines or patient tissue samples are used by the MNI algorithm to infer a model of regulatory interactions between genes (blue-filled circles indicate genes, arrows indicate regulatory influences). In phase 2, test condition expression data are filtered using the reconstructed network to distinguish the genes affected by a condition (red-filled circles) from the hundreds to thousands of genes that exhibit changes in expression. This procedure is applied to non-recurrent primary prostate cancer and metastatic prostate cancer samples to identify genetic mediators.

Abate-Shen and Shen, 2000; Feldman and Feldman, 2001; Navarro *et al*, 2002; Balaji *et al*, 2004; Shultz, 2005). Androgen independence is manifested in several ways. In some cases, sensitivity to low levels of androgen is increased by amplification, mutations and/or elevated levels or broadened specificity of co-activators of the androgen receptors. In other cases, activation of androgen receptors occurs in the absence of androgens due to crosstalk via other signaling pathways (Hobisch *et al*, 1995; Taplin *et al*, 1995; Abate-Shen and Shen, 2000; Feldman and Feldman, 2001; Navarro *et al*, 2002; Balaji *et al*, 2004; Shultz, 2005). It has been shown that almost all metastatic prostate cancers shift to an androgen-independent state (Abate-Shen and Shen, 2000; Navarro *et al*, 2002; Balaji *et al*, 2004). After anti-androgenic treatment, primary prostate cancers can also shift to an androgen-independent state and become recurrent (Abate-Shen and Shen, 2000; Navarro *et al*, 2002). In order to differentiate between the two groups optimally, we chose to analyze nine advanced-stage metastatic prostate cancer samples and 14 non-recurrent primary prostate cancer samples from LaTulippe *et al* (2002) (primary prostate cancer samples remained non-recurrent after a mean follow-up of 42 months).

The above points led us to hypothesize that the AR gene would be among the top genetic mediators identified by the MNI algorithm for the metastatic prostate cancer group only, indicative of its key role in androgen-independent metastatic prostate cancer. Moreover, because having amplifications, mutations and increased specificity for AR in androgen-independent prostate cancer raises the possibility that down-

stream genes in the AR pathway are also involved in the progression and metastasis of the disease, we further hypothesized that the AR signaling pathway would be highly enriched in the metastatic prostate cancer group.

Results and discussion

The list of the top 100 potential genetic mediators for non-recurrent primary and metastatic prostate cancer groups along with their expression change rankings is given in Supplementary Table 1.

The MNI algorithm identified the AR gene among the top genetic mediators in the metastatic prostate cancer group but not in the non-recurrent primary prostate group. We next subjected the 100 highest ranked genes in non-recurrent primary and metastatic prostate cancer groups to enrichment analysis for the AR signaling pathway. We found that the list of the top 100 genes for the metastatic prostate cancer was enriched ($P=1.5 \times 10^{-8}$) for the AR signaling pathway, in contrast to that for non-recurrent primary prostate cancer, which was not enriched (Figure 2). These results, which are consistent with our hypotheses, imply that the AR gene and the AR pathway are mediators of prostate cancer progression and metastasis.

Figure 2 includes the 20 genes transcriptionally regulated in the AR signaling pathway, identified among the top mediators for the metastatic prostate cancer group. It is thought that these genes may play a role in the acquisition of androgen-

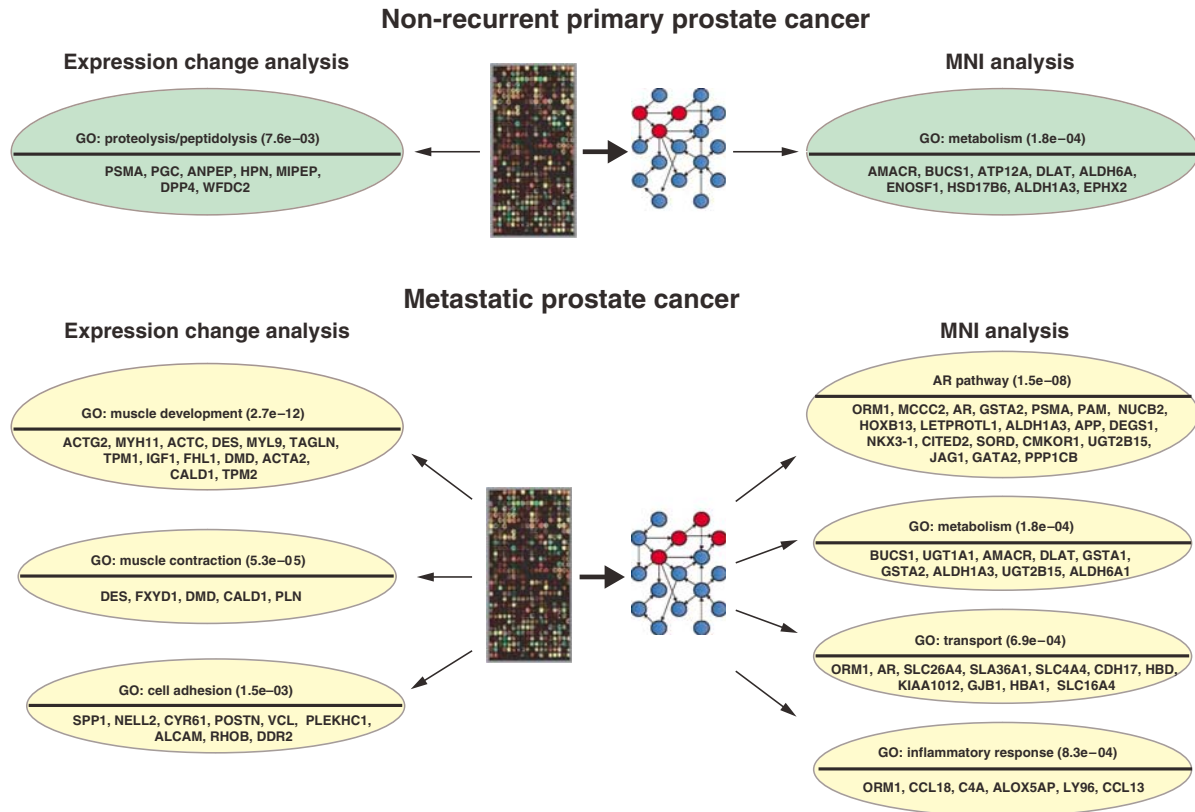


Figure 2 Significantly enriched pathways among top 100 genetic mediators identified by MNI and expression change alone, respectively, in non-recurrent primary and metastatic prostate cancer groups.

independent growth (Velasco *et al*, 2004). Among these 20 genes, six are well-known androgen-regulated genes (ARG)—AR, PSMA, HOXB13, NKX3-1, CITED2, UGT2B15 (Nelson *et al*, 2002; Velasco *et al*, 2004)—which have been shown to mediate metastatic disease progression. For example, PSMA (also known as FOLH1), prostate-specific membrane antigen 1, is used as a diagnostic and prognostic indicator for prostate cancer and is associated with prostate cancer aggressiveness and metastasis (Burger *et al*, 2002; Schmittgen *et al*, 2003; Kinoshita *et al*, 2005). HOXB13, homeobox B13, has been implicated in progression and metastasis in prostate cancer (Jung *et al*, 2004; Edwards *et al*, 2005; Zhao *et al*, 2005) and found to function as an AR repressor, modulating AR signaling (Jung *et al*, 2004). Loss of NKX3-1 expression, a homeodomain transcription factor, is found in 60–80% of human prostate carcinomas (Bhatia-Gaur *et al*, 1999) and has been associated with advanced prostate cancer and metastasis (Bowen, 2000). It has also been shown that a majority of *Nkx3.1*^{+/-}; *Pten*^{+/-} mice develop invasive adenocarcinoma in the lymph nodes (Abate-Shen *et al*, 2003). CITED2 is known to play a crucial role in the control of tumor hypoxia, which is associated with metastatic progression (Aprelikova *et al*, 2006). UGT2B15, UDP glucuronosyltransferase 2 family polypeptide B15, a steroid-metabolizing protein, has been found to be differentially expressed in androgen-independent bone marrow metastases following androgen ablation therapy (Guillemette *et al*, 1997; Hum *et al*, 1999; Stanbrough *et al*, 2006). We

discuss the clinical significance of some of the other top mediators in Supplementary information.

We next focused on GO-annotated pathways that were significantly overrepresented among the highly ranked genetic mediators. For our analysis, we subjected the 100 highest ranked genes identified by MNI in the metastatic and non-recurrent primary prostate cancer groups, respectively, to pathway analysis based on the GO biological process annotations obtained from Affymetrix.

The significantly enriched GO-annotated pathways for metastatic and non-recurrent primary prostate cancer based on MNI analysis are shown in Figure 2. These pathways fall into two categories, which are well-established processes identified as hallmarks of all cancers—metabolism and immune response (Weinberg and Hanahan, 2000). As malignant tumors consist of dividing cells, cancerous tissues have a high metabolic activity. The metabolism pathway is highly enriched in both non-recurrent primary and metastatic prostate cancer groups (Figure 2). Transport, indicative of high metabolism, is also significantly enriched in the metastatic prostate cancer group. The immune response is the main defense against cancer, and MNI successfully identifies the inflammatory response, the immune system-related biological pathway, in the metastatic prostate cancer group (Figure 2). Interestingly, some AR pathway genes are also part of the enriched GO-annotated pathways. Specifically, GSTA2, ALDH1A3 and UGTB15 in the metabolism pathway, ORM1 and AR

in the transport pathway and ORM1 in the inflammatory response are all androgen-responsive genes. This raises the possibility that the metabolism, transport and inflammatory response pathways in the metastatic prostate cancer group may be enriched as a result of increased activity of androgen-responsive genes in the AR signaling pathway.

In order to assess the relative merits of using MNI, a network-based approach, we performed pathway enrichment analysis on the top 100 ranked genes obtained by expression change alone and by using GSEA, a gene set enrichment tool (<http://www.broad.mit.edu/gsea/>). GO biological process pathway enrichment analysis, based on expression change alone, identified proteolysis and peptidolysis as a highly enriched pathway in the non-recurrent primary prostate cancer group and muscle development, muscle contraction and cell-adhesion pathways in the metastatic prostate cancer group (Figure 2). GO biological process pathway enrichment analysis using GSEA did not identify any significantly enriched pathways (Materials and methods). Compared to expression change alone, MNI failed to identify the cell-adhesion pathway, which is important in the spread and invasion of the cancer, as a highly enriched pathway. However, MNI was successful in eliminating false positives such as muscle contraction and muscle development pathways in the metastatic prostate cancer group. MNI's advantage was most apparent in predicting the AR signaling pathway as a mediator for metastatic prostate cancer. Neither expression change alone nor GSEA identified AR signaling as a highly enriched pathway. These analyses reveal the benefit of our network-based approach in that it highly ranks genes whose expression is not significantly different from normal and do not have coordinated expression, but which are relevant genetic mediators, such as the AR signaling pathway genes in the metastatic prostate cancer group.

Motivated by the above findings and the ability of our approach to differentiate between non-recurrent primary and

metastatic prostate cancer, we next applied the MNI algorithm to nine recurrent primary prostate cancer samples (which had recurred within a mean follow-up of 42 months) from LaTulippe *et al* (2002). We hypothesized that the MNI ranking of the AR gene would move up as an indication of the aggressiveness of the disease. Consistent with this hypothesis, MNI ranked the AR gene 970, 155 and 9 for the non-recurrent primary, recurrent primary and metastatic prostate cancer groups, respectively (Figure 3). This finding suggests that the AR gene, in the context of the reverse-engineered network, can be used as a marker for detecting the aggressiveness of primary prostate cancers. Interestingly, expression change alone ranked the AR gene 641, 668 and 207 in the respective groups (Figure 3), indicating that expression change alone is incapable of capturing the differential involvement of the AR gene in recurrent and non-recurrent primary prostate cancers.

In this study, we showed that a network biology approach that filters expression profiles through a reverse-engineered gene network can be used to identify the genetic mediators and mediating pathways of a disease. Specifically, we identified key genetic mediators and pathways that have been implicated in the initiation, advancement and invasion of prostate cancer. Our approach extends the utility of whole-genome expression profiling, and may be useful as a predictive tool for identifying novel genetic mediators for other cancers, such as breast cancer and leukemia. Network-based techniques (Gardner *et al*, 2003; Barabasi and Oltvai, 2004; Basso *et al*, 2005; Segal *et al*, 2005; Yeang *et al*, 2005) may also prove useful for providing biological insight into the etiology and progression of other diseases.

Materials and methods

Microarray data were collected from five publicly available databases: the NIH Gene Expression Omnibus (GSE349, GSE1431, GSE1923, GSE3960), Oncomine (Giordano_Adrenal, Nutt_Brain, Huang_Breast,

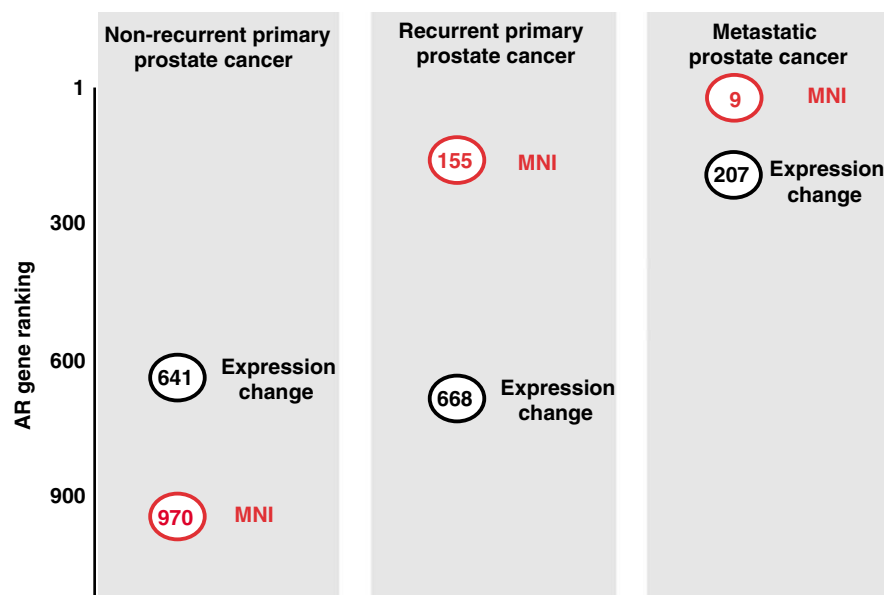


Figure 3 AR gene rankings based on MNI and expression change alone. The MNI ranking of the AR gene moves up as an indication of the aggressiveness of the disease. Expression change alone is unable to capture the differential involvement of the AR gene in the recurrent and non-recurrent primary prostate cancer groups.

LaTulippe_Prostate, Huang_Thyroid), EBI ArrayExpress (MEXP-441), Broad Institute Cancer Datasets (Bhattacharjee *et al*—Lung, Singh *et al*—Prostate) and the St Jude Research Datasets (Yeoh *et al*—Leukemia). The collected data were from experiments conducted on Affymetrix GeneChip Human Genome 95A or 95Av2; the combined microarrays have 12 600 overlapping probe sets. A total of 1144 experiments were collected based on 13 projects spanning seven different cancer types: adrenal, brain, breast, leukemia, lung, prostate and thyroid. CEL (cell intensity) files were processed individually for each project using RMAExpress. Each experiment in the data matrix was normalized by its mean to account for experimental variation between labs, and each probe set was normalized by its average across all experiments to obtain expression changes relative to a baseline. The MNI algorithm takes in as input the log-transformed expression ratios and standard errors. The data were log₂-transformed before being input to the MNI algorithm. As there were no repeated experiments in the collected set of microarray data, the standard error was set to 1.0 for all experiments and probe sets.

The MNI algorithm, which is described in detail in Supplementary information, takes as a training set all of the expression profiles except user-specified test profiles. We used as test profiles data from LaTulippe *et al* (2002), which includes samples from 14 non-recurrent primary prostate cancers, nine recurrent primary prostate cancers and nine metastatic prostate cancers located in the lymph node, bone, lung or soft tissue. The MNI algorithm was configured to output the top 200 mediators for each sample, together with the associated Z-scores for those probe sets. We set to zero the Z-score for probe sets that were not within the list of the top 200 probe sets identified as mediators for a given sample. To identify a characteristic list of genes within each group (i.e., non-recurrent primary, recurrent primary and metastatic prostate cancer), the Z-scores across samples and across probe sets for corresponding genes were averaged and ranked. The top 100 genes within that list were chosen to be reported as significant genetic mediators. A higher average Z-score is an indication of higher number of occurrences of a gene on the lists generated by the MNI algorithm in each group.

We compared the MNI rankings with those obtained using purely expression values. To make this comparison, we scored transcripts based on their differential expression from normal prostate tissue samples. Characteristic normal, non-recurrent primary, recurrent primary and metastatic expression profiles were created by averaging the normalized transcript expression for each of the experiments in those respective categories from LaTulippe *et al* (2002). The non-recurrent primary, recurrent primary and metastatic characteristic profiles were then divided by the normal profiles to obtain expression ratios. Ratios that were less than 1 were inverted to equally weight up- and downregulation. Expression rankings were obtained for each transcript using these two lists.

Pathway enrichment was performed on the top 100 genetic mediators identified for the non-recurrent primary and metastatic prostate cancer cases. The pathway annotations were based on the GO biological process annotations from Affymetrix for the HU95A and HU95Av2 chips. AR signaling pathway transcriptionally regulated genes were obtained from the NetPath database (<http://www.netpath.org/>). P-values for the pathway enrichment were calculated based on a hypergeometric distribution. We report significant pathway enrichments for groups with at least four members and P-values ≤ 0.01. GSEA (<http://www.broad.mit.edu/gsea/>) pathway enrichment was performed on the AR signaling pathway and all GO-annotated pathways. We used the GSEA-suggested threshold cutoff FDR=0.25.

Supplementary information

Supplementary information is available at the *Molecular Systems Biology* website (www.nature.com/msb).

Acknowledgements

This work was supported by the NSF, NIH and the National Heart, Lung and Blood Proteomics Initiative.

References

- Abate-Shen C, Banach-Petrosky WA, Sun X, Economides KD, Desai N, Gregg JP, Borowsky AD, Cardiff RD, Shen MM (2003) Nkx3.1; PTEN mutant mice develop invasive prostate adenocarcinoma and lymph node metastases. *Cancer Res* **63**: 3886–3890
- Abate-Shen C, Shen M (2000) Molecular genetics of prostate cancer. *Genes Dev* **14**: 2410–2434
- Aprelikova O, Wood M, Tackett S, Chandramouli GV, Barrett JC (2006) Role of ETS transcription factors in the hypoxia-inducible factor-2 target gene selection. *Cancer Res* **66**: 5641–5647
- Balaji KC, Rao PS, Smith DJ, Louis S, Smith LM, Sherman S, Bacich D, O’Keefe D (2004) Microarray analysis of differential gene expression in androgen independent prostate cancer using a metastatic human prostate cancer cell line model. *Urol Oncol* **22**: 313–320
- Barabasi AL, Oltvai ZN (2004) Network biology: understanding the cell’s functional organization. *Nat Rev Genet* **5**: 101–113
- Basso K, Margolin AA, Stolovitzky G, Klein U, Dalla-Favera R, Califano A (2005) Reverse engineering of regulatory networks in human B cells. *Nat Genet* **37**: 382–390
- Bhatia-Gaur R, Donjacour AA, Sciavolino PJ, Kim M, Desai N, Young P, Norton CR, Gridley T, Cardiff RD, Cunha GR, Abate-Shen C, Shen MM (1999) Roles for *Nkx3.1* in prostate development and cancer. *Genes Dev* **13**: 966–977
- Bowen C (2000) Loss of NKX3.1 expression in human prostate cancers correlates with tumor progression. *Cancer Res* **60**: 6111–6115
- Burger MJ, Tebay MA, Keith PA, Samaratunga HM, Clements J, Lavin MF, Gardiner RA (2002) Expression analysis of delta-catenin and prostate-specific membrane antigen: their potential as diagnostic markers for prostate cancer. *Int J Cancer* **100**: 228–237
- di Bernardo D, Thompson MJ, Gardner TS, Chobot SE, Eastwood EL, Wojtovich AP, Elliott SJ, Schaus SE, Collins JJ (2005) Chemogenomic profiling on a genome-wide scale using reverse-engineered gene networks. *Nat Biotechnol* **23**: 377–383
- Edwards S, Campbell C, Flohr P, Shipley J, Giddings I, Te-Poele R, Dodson A, Foster C, Clark J, Jhavar S, Kovacs G, Cooper CS (2005) Expression analysis onto microarrays of randomly selected cDNA clones highlights HOXB13 as a marker of human prostate cancer. *Br J Cancer* **92**: 376–381
- Feldman BJ, Feldman D (2001) The development of androgen-independent prostate cancer. *Nat Rev Cancer* **1**: 34–45
- Gardner TS, Bernardo D, Lorenz D, Collins JJ (2003) Inferring genetic networks and identifying compound mode of action via expression profiling. *Science* **301**: 102–105
- Guillemette C, Lévesque E, Beaulieu M, Turgeon D, Hum DW, Bélanger A (1997) Differential regulation of two uridine diphospho-glucuronosyltransferases, UGT2B15 and UGT2B17, in human prostate LNCaP cells. *Endocrinology* **138**: 2998–3005
- Hobisch A, Culig Z, Radmayr C, Bartsch G, Klocker H, Hittmair A (1995) Distant metastases from prostatic carcinoma express androgen receptor protein. *Cancer Res* **55**: 3068–3072
- Hum DW, Belanger A, Levesque E, Barbier O, Beaulieu M, Albert C, Vallee M, Guillemette C, Tchernof A, Turgeon D, Dubois S. (1999) Characterization of UDP-glucuronosyltransferases active on steroid hormones. *J Steroid Biochem Mol Biol* **69**: 413–423
- Jung C, Kim RS, Zhang HJ, Lee SJ, Jeng MH (2004) HOXB13 induces growth suppression of prostate cancer cells as a repressor of hormone-activated androgen receptor signaling. *Cancer Res* **64**: 9185–9192
- Kinoshita Y, Kuratsukuri K, Newman N, Rovito PM, Kaumaya PT, Wang CY, Haas GP (2005) Targeting epitopes in prostate-specific membrane antigen for antibody therapy of prostate cancer. *Prostate Cancer Prostatic Dis* **8**: 359–363
- LaTulippe E, Satagopan J, Smith A, Scher H, Scardino P, Reuter V, Gerald WL (2002) Comprehensive gene expression analysis of prostate cancer reveals distinct transcriptional programs associated with metastatic disease. *Cancer Res* **62**: 4499–4506

- Navarro D, Luzardo OP, Fernandez L, Chesa N, Diaz-Chico BN (2002) Transition to androgen independence in prostate cancer. *J Steroid Biochem Mol Biol* **81**: 191–201
- Nelson PS, Clegg N, Arnold H, Ferguson C, Bonham M, White J, Hood L, Lin B (2002) The program of androgen-responsive genes in neoplastic prostate epithelium. *Proc Natl Acad Sci USA* **99**: 11890–11895
- Schmittgen TD, Teske S, Vessella RL, True LD, Zakrajsek BA (2003) Expression of prostate specific membrane antigen and three alternatively spliced variants of PSMA in prostate cancer patients. *Int J Cancer* **107**: 323–329
- Segal E, Friedman N, Kaminski N, Regev A, Koller D (2005) From signatures to models: understanding cancer using microarrays. *Nat Genet* **37** Suppl: S38–S45
- Shultz WA (2005) *Molecular Biology of Human Cancers*. The Netherlands: Springer
- Stanbrough M, Bubley GJ, Ross K, Golub TR, Rubin MA, Penning TM, Febbo PG, Balk SP (2006) Increased expression of genes converting adrenal androgens to testosterone in androgen-independent prostate cancer. *Cancer Res* **66**: 2815–2825
- Taplin ME, Bubley GJ, Shuster TD, Frantz ME, Spooner AE, Ogata GK, Keer HN, Balk SP (1995) Mutation of the androgen-receptor gene in metastatic androgen-independent prostate cancer. *N Engl J Med* **332**: 1393–1398
- Velasco AM, Gillis KA, Li Y, Brown EL, Sadler TM, Achilleos M, Greenberger LM, Frost P, Bai W, Zhang Y (2004) Identification and validation of novel androgen-regulated genes in prostate cancer. *Endocrinology* **145**: 3913–3924
- Weinberg RA, Hanahan D (2000) The hallmarks of cancer. *Cell* **100**: 57–70
- Yeang CH, Mak HC, McCuine S, Workman C, Jaakkola T, Ideker T (2005) Validation and refinement of gene-regulatory pathways on a network of physical interactions. *Genome Biol* **6**: R62.1–R62.10
- Zhao Y, Yamashita T, Ishikawa M (2005) Regulation of tumor invasion by HOXB13 gene overexpressed in human endometrial cancer. *Oncol Rep* **13**: 721–726