JMB

Available online at www.sciencedirect.com

SCIENCE d DIRECT°

ELSEVIER

COMMUNICATION

Androgen Receptor Binding Sites Identified by a GREF_GATA Model

Katsuaki Masuda[1], Thomas Werner[2], Shilpi Maheshwari[1] Matthias Frisch[2], Soyon Oh[1], Gyorgy Petrovics[1], Klaus May[2] Vasantha Srikantan[1], Shiv Srivastava[1] and Albert Dobi[1]*

[1]*Center for Prostate Disease Research, Department of Surgery, Uniformed Services University, Rockville, MD 20852, USA*

[2]*Genomatix Software GmbH D-80339 Munich, Germany*

Changes in transcriptional regulation can be permissive for tumor progression by allowing for selective growth advantage of tumor cells. Tumor suppressors can effectively inhibit this process. The *PMEPA1* gene, a potent inhibitor of prostate cancer cell growth is an androgen-regulated gene. We addressed the question of whether or not androgen receptor can directly bind to specific *PMEPA1* promoter upstream sequences. To test this hypothesis we extended *in silico* prediction of androgen receptor binding sites by a modeling approach and verified the actual binding by *in vivo* chromatin immunoprecipitation assay. Promoter upstream sequences of highly androgen-inducible genes were examined from microarray data of prostate cancer cells for transcription factor binding sites (TFBSs). Results were analyzed to formulate a model for the description of specific androgen receptor binding site context in these sequences. *In silico* analysis and subsequent experimental verification of the selected sequences suggested that a model that combined a GREF and a GATA TFBS was sufficient for predicting a class of functional androgen receptor binding sites. The GREF matrix family represents androgen receptor, glucocorticoid receptor and progesterone receptor binding sites and the GATA matrix family represents GATA binding protein 1–6 binding sites. We assessed the regulatory sequences of the *PMEPA1* gene by comparing our model-based GREF_GATA predictions to weight matrix-based predictions. Androgen receptor binding to predicted promoter upstream sequences of the *PMEPA1* gene was confirmed by chromatin immunoprecipitation assay. Our results suggested that androgen receptor binding to cognate elements was consistent with the GREF_GATA model. In contrast, using only single GREF weight matrices resulted in additional matches, apparently false positives. Our findings indicate that complex models based on datasets selected by biological function can be superior predictors as they recognize TFBSs in their functional context.

© 2005 Elsevier Ltd. All rights reserved.

Keywords: model; androgen receptor; prostate cancer; promoter database; GATA factor

*Corresponding author

Alterations in transcription regulation are hallmarks of cancer progression. Decreases in the expression of specific genes that control cell proliferation may significantly contribute to malignant transformation. The *PMEPA1* gene has been recognized as a potent cell growth inhibitor of prostate cancer cells,[1] with decreased expression of *PMEPA1* associated with higher pathologic stage in prostate cancer specimens. Function for PMEPA1 in the ubiquitin-proteasome pathway is suggested by shared structural similarities between PMEPA1 and

---

Table 1. Androgen-inducible gene training set

Gene	Accession	Locus	Contig	Absolute positions (strand)
KLK3 (PSA)	HT2351	19q13.41	NM_145864	23,624,361–23,626,361 (+)
NDRG1	D87953	8q24.3	NT_008046	47,529,674–47,527,674 (−)
NKX3.1	U80669	8p21	NT_023666	1,916,763–1,914,763 (−)
IHPK1	D87452	3p21.31	NT_022517	49,711,601–49,709,601 (−)

Gene symbols, accession numbers and chromosomal locations (*Homo sapiens* NCBI Build 35, http://www.ncbi.nlm.nih.gov) are shown.

Nedd4-binding proteins. Interestingly, *PMEPA1* gene expression was shown to be regulated by androgens in prostate cancer cells.[2–4] In order to assess androgen receptor (AR) binding to regulatory elements of the *PMEPA1* gene we integrated computation biology with wetlab experimentation. Computation molecular biology of gene regulation is an emerging frontier of bioinformatics.[5] In the past, conclusive identification of functionally active transcription factor binding sites (TFBSs) relied mainly on laborious wetlab experimentations. However, recently chromatin immunoprecipitation (ChIP)-based assays such as "ChIP-chip" and ChIP Display were shown to be robust techniques to assess binding sites experimentally genome-wide.[6,7] In support to these efforts, robust platforms are currently available for *in vitro* selection of single TFBSs to assess optimal cognate sites for various elements.[8–10] Although bioinformatic tools are available for the prediction of TFBSs, we preferred a weight matrix approach.[11] However, the application of any weight matrix approach to long sequences is hampered by a large number of false positive predictions. Although examining the evolutionary conservation of transcription factor binding can enhance the specificity of *in silico* predictions,[12,13] this technique can be confounded by species-specific genes. Notably, many human cancer-related genes as illustrated by prostate specific antigen (PSA, *KLK3*) barely resemble their orthologs from other mammalian species.[14] Among mammals the pathogenesis of prostate cancer (CaP) is unique to man. The only mammal known to develop CaP in a similar fashion is the canine.[15] Fortunately, biological context can significantly enhance the quality of TFBS models.[16,17] Biologically linked sets of transcription factors observed in the combination of specific TFBSs can significantly improve prediction quality.[5] Consistent with this notion, we selected a training set of genes based on gene expression kinetics and dose–response of prostate cancer cells to androgen hormone stimulus.[4] Furthermore, we reasoned that gene regulation by the combination of closely situated binding sites is an essential feature of functional transcription regulatory regions.[18,19] Therefore, we formulated an androgen receptor-binding model from the combination of adjacent matrices. We experimentally verified this model and went on to assess the utility of the model by testing upstream sequences of the androgen-inducible gene, *PMEPA1.*[2]

Defining a model for androgen receptor binding in the context of prostate-specific hormone induction

Upstream sequences of highly androgen-inducible genes were selected as a training set from our previous oligonucleotide array experiments in order to formulate a model that can predict functional androgen receptor binding sites in the context of androgen-induced prostate cancer cells (Table 1). Four genes were selected that stood out of the list of androgen-inducible genes both by the robustness of gene expression response (KLK3 = 58× and NDRG1 = 13×) and by kinetic considerations (IHPK1 = 9× in 1 h and NKX3.1 = 5× in 6 h).[4]

Extraction of 5′ upstream sequences relative to the transcription start sites of training set genes between positions −2000 and +1 was accomplished by the ElDorado system (Genomatix, Munich)† that is based on the genome assemblies from NCBI (*Homo sapiens* NCBI Build 35‡). Sequences were verified by database searches in the human genome. In this process upstream sequences of *KLK3* and *KLK2* genes were found to show high degrees of identity (approx. 80% within −1000; +1). Although, we previously found that *KLK2* was a highly androgen-inducible gene, *KLK2* was excluded from the training set to avoid sequence redundancy that may introduce bias into the model prediction. The training set was analyzed by the MatInspector[20] to identify potential androgen receptor binding sites. Androgen receptor can recognize DNA sequences represented by the glucocorticoid responsive element matrix family (GREF) that includes matrices of androgen receptor, glucocorticoid receptor and progesterone receptor binding sites. Two GREFs in *KLK3* were identified, one in *NDRG1*, two in *NKX3.1* and one within the *IHPK1* upstream sequences. In addition to the GREF family the program found 27 other matrix families to match within the sequences. The relation of each predicted site was assessed relative to the GREF positions. The distance criterion for the assessment was 20–33 base-pairs between the centers of two adjacent sites (within 2–3 DNA helical turn distance). Only the GATA matrix family matches showed distance correlation to the GREF

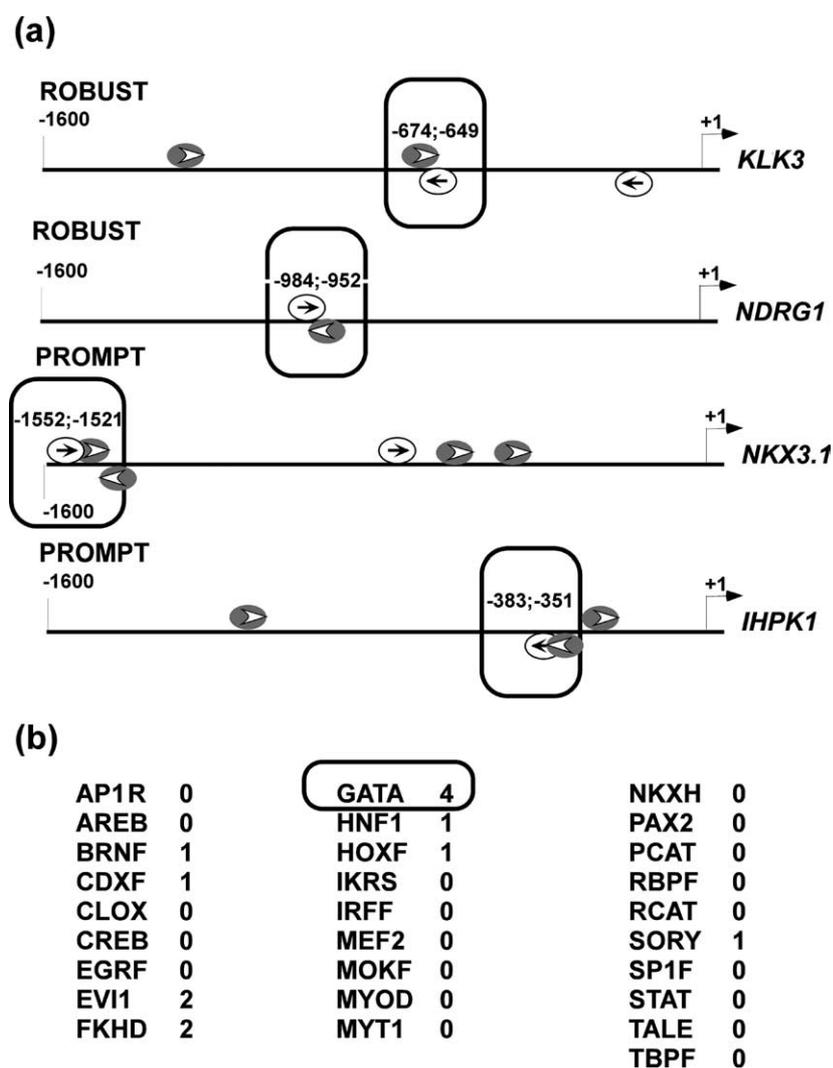† http://www.genomatix.de
‡ http://www.ncbi.nlm.nih.gov

Figure 1. Matrices within the training set sequences. (a) Position of matches of GREF (encircled filled arrows) and GATA (open arrowheads) matrix families within the training set of sequences. Arrowheads indicate the strand orientation. Adjacent GREF and GATA matrices (min. 20 and max. 33 bp interval between the centers of matrices) and the corresponding matrix anchor positions are boxed. (b) List of all matches of matrix families identified within the training set sequences. Numbers indicate how often a given matrix matched within the pre-defined range of a GREF matrix match (20–33 bp). Only GATA matrix family matches fulfilled this condition in all four sequences.

matrix family matches (Figure 1). Indeed, the identification of potential GATA binding sites close to androgen responsive elements is consistent with previous findings that suggested critical roles for GATA2 and GATA3 transcription factors in androgen receptor functions.[21] It has been demonstrated that GATA-driven expression of SV40 T-antigen resulted in androgen-independent prostate cancer in transgenic mice.[22] Moreover, the *KLK3* gene upstream androgen responsive enhancer (AREIII) (Figure 2) that is embedded in a cluster of adjacent GATA factor binding sites was shown to be essential for optimal androgen induction of *KLK3* gene expression.[22]

Experimental validation of the GREF_GATA model

By assessing various core similarities we found that a threshold of 0.85 core similarity was optimal for both matrices. Matrix similarities were set to default values (optimized). The model was then defined using the FastM module of GEMS Launcher[13] and was successfully checked for correctness by searching the sequences of the training data set by ModelInspector.[23] To compensate for model overfitting we assessed the validity of our findings by an independent experimental approach. For that purpose we tested the predicted sites in our training set for androgen receptor binding in androgen-induced prostate cancer cells by *in vivo* ChIP. The *KLK3* gene androgen responsive enhancer (ARE III) was assayed as an additional positive control (Figure 2). Hormone-induced androgen receptor binding to predicted sites was evident in four out of the five tested gene promoter upstream sequences.

Androgen receptor binds preferentially to the region defined by the GREF_GATA model within the *PMEPA1* gene promoter upstream sequences in androgen-induced LNCaP cells

We used the experimentally supported GREF_GATA model to analyze the upstream −8000;+1 sequences of the androgen-inducible gene,
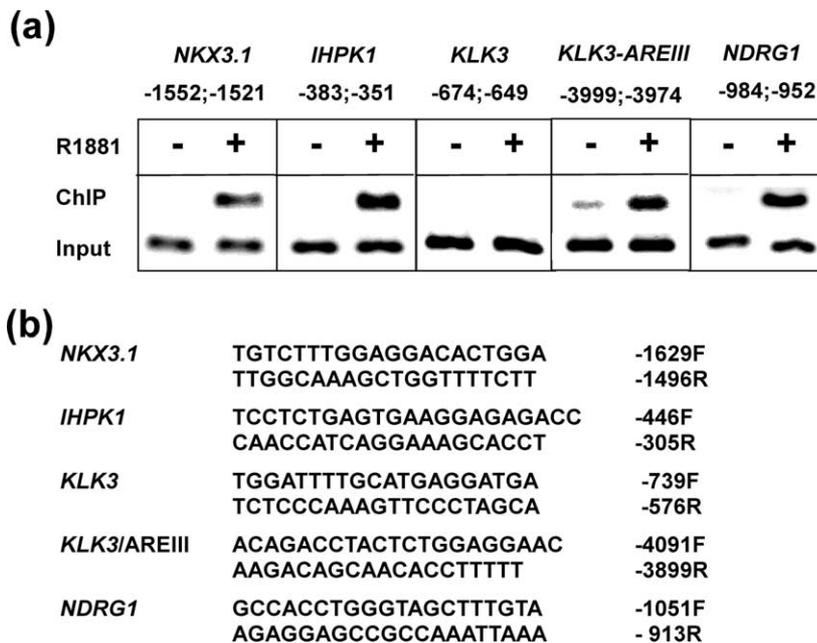
Figure 2. Experimental verification of predicted androgen receptor binding sites. (a) ChIP assay results of androgen-inducible gene promoters are shown. GREF_GATA model-match positions are indicated under gene symbols. For ChIP assay, prostate cancer cells (LNCaP, American Type Culture Collection, Rockville, MD) were cultured in charcoal treated androgen-free fetal bovine serum (Invitrogen, Carlsbad, CA) for five days to deplete androgens from the cells. After androgen depletion cell cultures were incubated in the presence or absence of 3.3 nM of R1881 non-metabolizable androgen (DuPont, Boston, MA) for 24 h. For ChIP assay the medium was removed from the cells by aspiration. In each parallel, 2×10^6 LNCaP cells were incubated in PBS buffer with 1% formaldehyde (Sigma, St. Louis, MO) at 0 °C for 10 min followed by incubation at 37 °C for 15 min. The crosslinking reaction was stopped by the addition of glycine to a final concentration of 0.125 M and the cells were washed three times with ice-cold 1xPBS in the presence of Complete Mini protease inhibitor (Roche, Indianapolis, IN) and Phosphatase Inhibitor Cocktail I and II (Sigma, St. Louis, MO). Inhibitors were added to the reaction mixtures in all subsequent steps. The cells were lysed (50 mM Tris–HCl (pH 8.0), 1% (w/v) SDS, 10 mM EDTA) and the lysates were incubated at 0 °C for 10 min. The chromatin was fragmented by sonicating the lysates with a VirSonic 100 sonicator microtip (VIRTIS, Gardiner, NY) applying four times 10 s bursts at energy level 4, followed by 50 s of cooling after each burst. The cell debris were separated by centrifugation at 10,000g at 4 °C in a microcentrifuge and the supernatants were diluted ten times (16.7 mM Tris–HCl (pH 8.0), 167 mM NaCl, 1.1% TritonX-100, 1.2 mM Na-EDTA, 0.01% SDS). The suspensions were filtrated through 0.45 μm filter syringes and the flow-through was collected in pre-chilled tubes. Aliquots of 50 μl of 50% immobilized and pre-blocked protein A/G-bead slurry (PIERCE, Rockford, IL) were added to an aliquot of 250 μl of the filtrates and the sample tubes were agitated on a head-to-head rotator for 1 h at 4 °C. Subsequently the beads were removed by centrifugation (4 °C, 3000g, 2 min) and 200 μl of the clear supernatants were transferred into pre-chilled microcentrifuge tubes. To each tube, 2 μl (0.33 μg/μl) of anti-androgen receptor antibody AR(PG21) (Upstate, Lake Placid, NY) was added. The immunoprecipitation reactions were incubated for 12 h at 4 °C with gentle head-to-head rotation. Subsequently, the chromatin–antibody complexes were captured by the addition of 60 μl of immobilized protein A/G slurry by incubating the suspensions at 4 °C (1 h) continuing gentle head-to-head rotation. After incubation the beads were collected by brief centrifugation and the beads were washed four times with ice-cold buffer of 20 mM Tris–HCl (pH 8.0), 150 mM NaCl, 2 mM Na-EDTA, 0.1% SDS and 1% TritonX-100, once at 24 °C with the buffer of 20 mM Tris–HCl (pH 8.0), 250 mM LiCl, 1 mM Na-EDTA, 0.5% IGEPAL (Sigma, St. Louis, MO) non-ionic detergent and 0.5% sodium-deoxycholate (Sigma, St. Louis, MO) and twice with the solution of 10 mM Tris–HCl (pH 8.0), 1 mM Na-EDTA at room temperature. After each washing step the beads were collected by centrifugation (24 °C, 1 min, 4000g). The beads were resuspended in TE-buffer (10 mM Tris–HCl (pH 7.5), 1 mM Na-EDTA) containing 0.5% of SDS and Proteinase K enzyme (Roche, Indianapolis, IN) to a final concentration of 0.5 mg/ml to digest the protein components. The reaction mixtures were incubated for 12 h at 37 °C. The clear supernatants were recovered by spinning the sample in a microcentrifuge and the suspensions were incubated at 65 °C for 6 h to revert the chemical crosslinking. Then the samples were purified by phenol/chloroform extraction and subsequent precipitation with ethanol. The pellets were dissolved in TE buffer. The input genomic DNA samples were treated in the same way except that no immunoprecipitation was performed. (b) Specific sets of primer pairs were selected for PCR reactions in order to amplify the regions of predicted GREF sequences within the promoter upstream sequences of *KLK3, NDRG1, NKX3.1* and *IHPK1* genes. The amplification was carried out in a T-Gradient Thermoblock (Biometra, Tampa, FL) (94 °C, 15 s; annealing at 58 °C (for *KLK3*/AREIII 55 °C); 30 s, 72 °C, 1 min). To achieve linear amplification ranges ChIP and input products were amplified with 43 and 38 cycles, respectively.

PMEPA1. The 5′ distal boundary of the selected region (−8000) excluded a cluster of Alu repeats and a matrix attachment region (S/MAR) and the proximal boundary was extended until the transcription initiation site (+1). Two model matches at positions −2134 (GREF)/−2156 (GATA) and −230 (GREF)/−250 (GATA) were found with model scores of 95.1% and 92.6%, respectively (Figure 3(a)). In contrast, eight GREF matrix family matches were found that included preferred sites of androgen, progesterone and glucocorticoid-responsive elements (ARE, PRE, and GRE, respectively). *In vivo* binding of androgen receptor to the *PMEPA1* gene upstream sequences in LNCaP
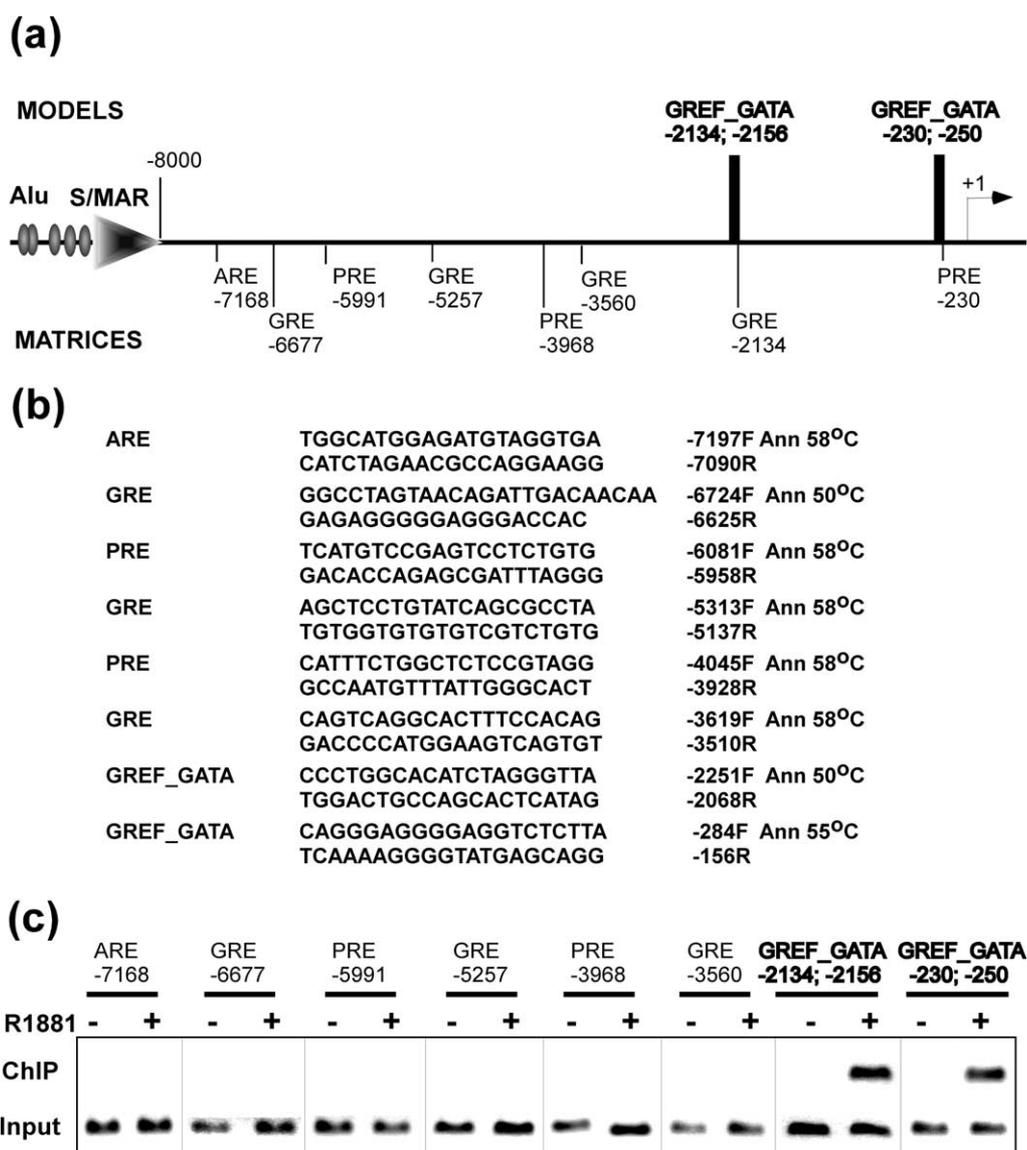
Figure 3. Positions of model and matrix-matches and experimental validation of corresponding androgen receptor binding sites within the *PMEPA1* gene upstream sequences. (a) GREF_GATA model-matches are shown in bold. Predicted matches for GREF matrix family members ARE, GRE, and PRE, respectively, are shown below. (b) Sequences of forward (F) and reverse (R) primer pairs and corresponding annealing temperatures are listed. (c) ChIP results with anti-androgen receptor antibodies and the control input amplification products (Input) are shown. Prostate cancer cells (LNCaP) were incubated in the absence (−) or presence (+) of synthetic androgen (R1881). For details see the legend to Figure 2.

prostate cancer cells in response to R1881 synthetic androgen stimulus was examined. Androgen receptor binding was tested by ChIP assay using anti-AR antibody and primer sets to amplify the predicted binding sites of GREF matrix matches within the −8000;+1 region of the *PMEPA1* gene (Figure 3(b)). Our GREF_GATA model correctly predicted binding of androgen receptor to its cognate elements in response to androgen induction (Figure 3(c)). It was surprising that a seemingly excellent GREF/ARE matrix match (−7168 in Figure 3) was not a functional androgen receptor binding site as opposed to the two predicted

GREF_GATA models, both of which could be validated by ChIP. It is remarkable that both sites were recognized as glucocorticoid and progesterone receptor binding sites, demonstrating that the matrix family concept is valid (as shown by the chromatin immunoprecipitation) as well as essential for model definition.

Promoter database search with the GREF_GATA model

The model has shown good predictive power in the case of the *PMEPA1* gene. Therefore, the entire

Genomatix Promoter Database (GPD) was searched for promoters from human annotated genes with the GREF_GATA model (ModelInspector, default settings). The search identified 995 matches (belonging to 933 different genes) out of 36,943 promoters analyzed, a match rate of 2.7%, indicating moderate selectivity. Statistical analysis of the 933 genes revealed that the selection of identified genes was not random and the gene ontology (GO) category "sensory perception of chemical stimulus" was clearly over-represented in the match list. This most significant group in the ontology "biological process" (z-score 6.19) contained nine genes, seven of which were olfactory receptors. Within the ontology "molecular function" the group "olfactory receptor activity" was the top-scoring group (z-score 7.09) containing more than five genes. Further inspection of the output revealed 18 genes out of a total of 270 olfactory receptor genes in the human genome (numbers taken from ElDorado). Thus olfactory receptors were 2.5-fold more often found than expected at random. Promoters of nine genes of the top-scoring group were aligned using the program DiAlign to exclude the possibility that observed matches were due to extensive sequence similarities due to recent gene/promoter duplication.[24] The resulting sequence similarities were (except for the pair OR5I1/OR6C3, 32%) all below the random similarity of 25%, strongly suggesting that the conservation of the promoter framework was due to functional constraints rather than sequence similarity. Assessment of specificity and sensitivity would require knowledge of the absolute number of true positive matches. This obstacle was circumvented by assessing whether the match list was enriched in potentially androgen receptor associated genes or not, an approach feasible based on available data. The total number of genes co-cited with androgen receptor in the literature is 1034. Co-citation analysis was done with BiblioSpere†, a literature-mining tool based on PubMed abstracts‡. A total of 51 of the 933 genes identified by the GREF_GATA model were co-cited with androgen receptor. To determine if these results indicate a significant association with androgen receptor, 30 random picks of genes were identified in the database. The number of genes co-cited with androgen receptor in the literature within the random group was determined. A z-score (number of genes co-cited with AR in model search–average number of genes co-cited with AR in random picks) / standard deviation of random pick results was calculated. The resulting z-score was 4.4, indicating a significant enrichment of androgen receptor co-cited genes in the model-search result.

For model validation we assessed the binding of androgen receptor to the promoter sequences of genes within the prostate cancer biological context by *in vivo* ChIP assay. The assay indicated androgen receptor binding to six out of eight tested gene promoter elements in response to androgen (R1881) stimulus in LNCaP cells (Figure 4). These data suggested that the GREF_GATA model was a powerful predictor of biologically active androgen receptor binding sites.

## Biological context is an important factor in training set selection

Androgens and their receptor are the main therapeutic targets in prostate cancer.[25] While androgen blockade has been a significant therapeutic choice,[26] failure of androgen-blockade therapy can dramatically reduce the survival rate.[27] Identification of specific AR cognate binding sequences throughout the entire human genome relying exclusively on experimental high throughput TFBS identification techniques may not be practical. A high rate of missed TFBS in robust scanning techniques has been reported.[28] On the other hand, bioinformatic tools are readily available to predict potential androgen receptor binding sites. However, all single binding site oriented approaches are plagued by an enormous proportion of presumably false positive predictions, referred to as the "futility theorem".[5] Therefore, we addressed the question of whether or not androgen receptor binding sites can be more selectively predicted by a model-based approach. We found that a model composed of glucocorticoid responsive element matrices (GREF) adjacent to GATA matrices (GATA) allowed the identification of a novel class of biologically active androgen receptor binding sites. Selection of a training set of promoters from a defined biological context was crucial for successful model generation.[29,30] A genome-wide promoter search with the GREF_GATA model has identified 933 genes, 51 of which were found to be co-cited with androgen receptor in the literature. Five examples of known androgen/prostate cancer related genes are: six transmemebrane epithelial antigen of prostate 2, *STEAP2* (Loc261729);[31] acid phosphatase of prostate, *ACPP* (Loc55);[32] breast cancer 2 early onset, *BRCA2* (Loc 675);[33] aconitase 2 mitocondrial, *ACO2* (Loc50);[34] androgen-induced proliferation inhibitor, *APRIN* (Loc23047).[35] It is interesting to note that the top scoring GO-category of olfactory receptors was functionally linked to androgen receptor in a previous study.[36] These findings are consistent with the known role that androgen receptor plays in gender-specific sensory functions.[37–39]

In summary, we selected genes based on their expression response to androgen hormone induction before defining our training set. The hormone-responsive prostate cancer cell line LNCaP as experimental host was selected for its wide use as an experimental model.[40] Our previous oligonucleotide array expression platform directly

---

(a)
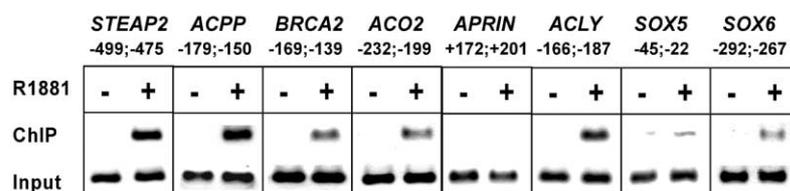
(b)

STEAP2	TCCTCCTATCCGAGGGAGTT	-560F
	TGGTTCTCATCTCCCCAGTC	-382R
ACPP	ACAGCTTTCCCCGCTCTT	-203F
	GAAGGAGGAGCCACACTCAT	-53R
BRCA2	GGGTCAGCGAGAAGAGAACA	-190F
	TGCGGGTATTTCTCAGTGTG	-80R
ACO2	GATTTGGATGGCGGAGATAA	-258F
	AGACACGCAAGAGGAAATCG	-150R
APRIN	GCAGAATTCCTTTCTGCTTTG	+107F
	TGTGACTGAGAACATACCATAACC	+216R
ACLY	AGTGCTCCTGTGGGTTTGTT	-250F
	CAGGAGGTACTGGGGAGTGA	-137R
SOX5	TTGGTAGAGTAATTTGATGGGAAA	-107F
	TGATCATTAGAACATGAGCTGTG	- 7R
SOX6	AAGCCTGTTTCTTTCTTTCTCTT	-330F
	ATGCCACATTTCCATCCAAT	-210R

Figure 4. Hormone-induced androgen receptor binding to predicted binding sites in the genome of prostate cancer cells. (a) ChIP assay of *STEAP2*, *ACPP BRCA2*, *ACO2*, *APRIN*, *ACLY*, *SOX5* and *SOX6* gene promoters. GREF_-GATA model-match positions are shown under gene symbols. Model-match was found downstream relative to an alternative transcription initiation site of the *APRIN* gene. LNCaP cells were incubated in the presence (+) or absence (−) of R1881 synthetic androgen. Input marks the amplification product of control genomic DNA. *In vivo* ChIP assay was carried out as described in the legend to Figure 2. (b) Primer sequences and positions relative to the transcription initiation sites of the tested genes are listed. In the amplification step all primers were annealed at 58 °C.

revealed the time kinetic- and dose-dependent responses of gene expression profiles. The *KLK3*, *NDRG1*, *NKX3.1*, and *IHPK1* genes were selected as our training set based on this biological evidence. We have defined GREF_GATA, a novel model to predict androgen responsive elements that will provide new insights into the androgen-regulated transcriptome in prostate cancer. Identifying direct targets of AR will allow the detection of critical changes of androgen signaling in prostate cancer progression.

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2005.09.009

References

1. Xu, L. L., Shi, Y., Petrovics, G., Sun, C., Makarem, M., Zhang, W. *et al*. (2003). PMEPA1, an androgen-regulated NEDD4-binding protein, exhibits cell growth inhibitory function and decreased expression during prostate cancer progression. *Cancer Res.* **63**, 4299–4304.
2. Xu, L. L., Shanmugam, N., Segawa, T., Sesterhenn, I. A., McLeod, D. G., Moul, J. W. *et al*. (2000). A novel androgen-regulated gene, PMEPA1, located on chromosome 20q13 exhibits high level expression in prostate. *Genomics*, **66**, 257–263.
3. Xu, L. L., Su, Y. P., Labiche, R., Segawa, T., Shanmugam, N., McLeod, D. G. *et al*. (2001). Quantitative expression profile of androgen-regulated genes in prostate cancer cells and identification of prostate-specific genes. *Int. J. Cancer*, **92**, 322–328.
4. Segawa, T., Nau, M. E., Xu, L. L., Chilukuri, R. N., Makarem, M., Zhang, W. *et al*. (2002). Androgen-

induced expression of endoplasmic reticulum (ER) stress response genes in prostate cancer cells. *Oncogene*, **21**, 8749–8758.

5. Wasserman, W. W. & Sandelin, A. (2004). Applied bioinformatics for the identification of regulatory elements. *Nature Rev. Genet.* **5**, 276–287.

6. Iyer, V. R., Horak, C. E., Scafe, C. S., Botstein, D., Snyder, M. & Brown, P. O. (2001). Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. *Nature*, **409**, 533–538.

7. Barski, A. & Frenkel, B. (2004). ChIP Display: novel method for identification of genomic targets of transcription factors. *Nucl. Acids Res.* **32**, e104.

8. Bulyk, M. L., Gentalen, E., Lockhart, D. J. & Church, G. M. (1999). Quantifying DNA-protein interactions by double-stranded DNA arrays. *Nature Biotechnol.* **17**, 573–577.

9. Roulet, E., Busso, S., Camargo, A. A., Simpson, A. J., Mermod, N. & Bucher, P. (2002). High-throughput SELEX SAGE method for quantitative modeling of transcription-factor binding sites. *Nature Biotechnol.* **20**, 831–835.

10. Linnell, J., Mott, R., Field, S., Kwiatkowski, D. P., Ragoussis, J. & Udalova, I. A. (2004). Quantitative high-throughput analysis of transcription factor binding specificities. *Nucl. Acids Res.* **32**, e44.

11. Cartharius, K., Frech, K., Grote, K., Klocke, B., Haltmeier, M., Klingenhoff, A. *et al*. (2005). MatInspector and beyond: promoter analysis based on transcription factor binding sites. *Bioinformatics*, **21**, 2933–2942.

12. Dieterich, C., Rahmann, S. & Vingron, M. (2004). Functional inference from non-random distributions of conserved predicted transcription factor binding sites. *Bioinformatics*, **20**, I109–I115.

13. Klingenhoff, A., Frech, K., Quandt, K. & Werner, T. (1999). Functional promoter modules can be detected by formal models independent of overall nucleotide sequence similarity. *Bioinformatics*, **15**, 180–186.

14. Olsson, A. Y., Lilja, H. & Lundwall, A. (2004). Taxon-specific evolution of glandular kallikrein genes and identification of a progenitor of prostate-specific antigen. *Genomics*, **84**, 147–156.

15. Russell, P. J. & Voeks, D. J. (2003). Animal models of prostate cancer. *Methods Mol. Med.* **81**, 89–112.

16. Fessele, S., Maier, H., Zischek, C., Nelson, P. J. & Werner, T. (2002). Regulatory context is a crucial part of gene function. *Trends Genet.* **18**, 60–63.

17. Werner, T., Fessele, S., Maier, H. & Nelson, P. J. (2003). Computer modeling of promoter organization as a tool to study transcriptional coregulation. *FASEB J.* **17**, 1228–1237.

18. Kel-Margoulis, O. V., Romashchenko, A. G., Kolchanov, N. A., Wingender, E. & Kel, A. E. (2000). COMPEL: a database on composite regulatory elements providing combinatorial transcriptional regulation. *Nucl. Acids Res.* **28**, 311–315.

19. Werner, T. (2001). The promoter connection. *Nature Genet.* **29**, 105–106.

20. Quandt, K., Frech, K., Karas, H., Wingender, E. & Werner, T. (1995). MatInd and MatInspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data. *Nucl. Acids Res.* **23**, 4878–4884.

21. van den Bemd, G. J., Jhamai, M., Brinkmann, A. O. & Chang, G. T. (2003). The atypical GATA protein TRPS1

repress androgen-induced prostate-specific antigen expression in LNCaP prostate cancer cells. *Biochem. Biophys. Res. Commun.* **312**, 578–584.

22. Perez-Stable, C. M., Pozas, A. & Roos, B. A. (2000). A role for GATA transcription factors in the androgen regulation of the prostate-specific antigen gene enhancer. *Mol. Cell. Endocrinol.* **167**, 43–53.

23. Morgenstern, B., Dress, A. & Werner, T. (1996). Multiple DNA and protein sequence alignment based on segment-to-segment comparison. *Proc. Natl Acad. Sci. USA*, **93**, 12098–12103.

24. Morgenstern, B. (2004). DIALIGN: multiple DNA and protein sequence alignment at BiBiServ. *Nucl. Acids Res.* **32**, W33–W36.

25. Chen, C. D., Welsbie, D. S., Tran, C., Baek, S. H., Chen, R., Vessella, R. *et al.* (2004). Molecular determinants of resistance to antiandrogen therapy. *Nature Med.* **10**, 33–39.

26. Moul, J. W. (1998). Contemporary hormonal management of advanced prostate cancer. *Oncology (Huntingt)*, **12**, 499–508.

27. Huang, H. & Tindall, D. J. (2002). The role of the androgen receptor in prostate cancer. *Crit. Rev. Eukaryot. Gene Expr.* **12**, 193–207.

28. Buck, M. J. & Lieb, J. D. (2004). ChIP-chip: considerations for the design, analysis, and application of genome-wide chromatin immunoprecipitation experiments. *Genomics*, **83**, 349–360.

29. Krivan, W. & Wasserman, W. W. (2001). A predictive model for regulatory sequences directing liver-specific transcription. *Genome Res.* **11**, 1559–1566.

30. Liu, R., McEachin, R. C. & States, D. J. (2003). Computationally identifying novel NF-kappa B-regulated immune genes in the human genome. *Genome Res.* **13**, 654–661.

31. Porkka, K. P., Helenius, M. A. & Visakorpi, T. (2002). Cloning and characterization of a novel six-trans-membrane protein STEAP2, expressed in normal and malignant prostate. *Lab. Invest.* **82**, 1573–1582.

32. Winqvist, R., Virkkunen, P., Grzeschik, K. H. & Vihko, P. (1989). Chromosomal localization to 3q21–qter and two TaqI RFLPs of the human prostate-specific acid phosphatase gene (ACPP). *Cytogenet. Cell Genet.* **52**, 68–71.

33. Simard, J., Dumont, M., Labuda, D., Sinnett, D., Meloche, C., El-Alfy, M. *et al.* (2003). Prostate cancer susceptibility genes: lessons learned and challenges posed. *Endocr. Relat. Cancer*, **10**, 225–259.

34. Juang, H. H. (2004). Nitroprusside stimulates mitochondrial aconitase gene expression through the cyclic adenosine $3',5'$-monophosphate signal transduction pathway in human prostate carcinoma cells. *Prostate*, **61**, 92–102.

35. Geck, P., Szelei, J., Jimenez, J., Sonnenschein, C. & Soto, A. M. (1999). Early gene expression during androgen-induced inhibition of proliferation of prostate cancer cells: a new suppressor candidate on chromosome 13, in the BRCA2-Rb1 locus. *J. Steroid. Biochem. Mol. Biol.* **68**, 41–50.

36. Paredes, R. G., Lopez, M. E. & Baum, M. J. (1998). Testosterone augments neuronal Fos responses to estrous odors throughout the vomeronasal projection pathway of gonadectomized male and female rats. *Horm. Behav.* **33**, 48–57.

37. Cooke, B. M., Breedlove, S. M. & Jordan, C. L. (2003). Both estrogen receptors and androgen receptors contribute to testosterone-induced changes in the morphology of the medial amygdala and sexual arousal in male rats. *Horm. Behav.* **43**, 336–346.

38. Zakon, H. (2004). Physiology. Heeding the hormonal call. *Science*, **305**, 349–350.

39. Kritzer, M. (2004). The distribution of immunoreactivity for intracellular androgen receptors in the cerebral cortex of hormonally intact adult male and female rats: localization in pyramidal neurons making corticocortical connections. *Cereb. Cortex*, **14**, 268–280.

40. Horoszewicz, J. S., Leong, S. S., Chu, T. M., Wajsman, Z. L., Friedman, M., Papsidero, L. *et al*. (1980). The LNCaP cell line—a new model for studies on human prostatic carcinoma. *Prog. Clin. Biol. Res.* **37**, 115–132.

Edited by M. Yaniv