

Identification of novel steroid target genes through the combination of bioinformatics and functional analysis of hormone response elements

Kuniko Horie-Inoue^a, Kenichi Takayama^{a,b}, Hidemasa U. Bono^c, Yasuyoshi Ouchi^b,
Yasushi Okazaki^c, Satoshi Inoue^{a,b,*}

^a Division of Gene Regulation and Signal Transduction, Research Center for Genomic Medicine, Saitama Medical School, 1397-1 Yamane, Hidaka-shi, Saitama 350-1241, Japan

^b Department of Geriatric Medicine, Graduate School of Medicine, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8655, Japan

^c Division of Functional Genomics and Systems Medicine, Research Center for Genomic Medicine, Saitama Medical School, 1397-1 Yamane, Hidaka-shi, Saitama 350-1241, Japan

Received 18 October 2005

Available online 8 November 2005

Abstract

Steroid hormone receptors including androgen receptor (AR), glucocorticoid receptor (GR), progesterone receptor (PR), and mineralocorticoid receptor (MR) recognize and bind to identical consensus hormone response elements (HREs), which consist of two hexameric half-sites (5'-AGAACA-3') arranged as inverted repeats with a 3-bp spacer. Although only a few near-consensus HRE sequences have been identified in the transcriptional regulatory regions of known steroid target genes, it has been unclear whether the exact consensus sequences function as bona fide HREs in vivo. A genome-wide in silico screening of palindromic HREs identified 565 exact consensus sequences in human genome (NCBI 35 assembly). In this study, of 565 exact consensus elements, functional in vivo receptor binding was evaluated regarding 26 sequences located within 10 kb upstream to the 5' end of annotated genes through chromatin immunoprecipitation (ChIP) assay using cells endogenously expressing steroid hormone receptors. Hormone responsiveness of proximal gene expression was examined through quantitative RT-PCR. As far as performing ChIP assay for AR, GR, and PR, 14 of 26 elements significantly recruited at least one of the receptors by hormone treatment (>2-fold enrichment versus vehicle). In terms of gene expression in the vicinity of the above 14 functional perfect HREs, four genes were upregulated by >2-fold with hormone treatment. The present data suggest that the combination of bioinformatics analysis and quantitative experimental evaluation is useful to identify novel functional HREs that may contribute to the transcriptional regulation of steroid target genes.

© 2005 Elsevier Inc. All rights reserved.

Keywords: Androgen receptor; Progesterone receptor; Glucocorticoid receptor; Steroid target gene; Hormone response element; Chromatin immunoprecipitation; Quantitative PCR; Transcriptional start site

Steroid hormone receptors are nuclear receptors that play essential roles in various physiological and pathophysiological states. Forming complexes with coactivators and general transcription factors, ligand-stimulated steroid hormone receptors including androgen receptor (AR), glucocorticoid receptor (GR), progesterone receptor (PR), and mineralocorticoid receptor (MR) recognize and bind to hormone response elements (HREs) in the regulatory

regions of various hormone responsive genes, leading to the modulation of target gene transcription [1]. Despite the degeneracy of the nucleotide sequence of the HREs, steroid hormone receptors generate distinct hormone-specific responses. The palindromic 15-bp sequence (5'-AGAACA_nTTGTTCT-3') that contains inverted repeats with a 3-bp spacer (IR3 sequence) has been identified as the consensus sequence for HRE [2], though only a few perfect consensus HREs have yet been identified in the regulatory regions of known hormone responsive genes. A question that continues to engage the steroid receptor field is

* Corresponding author. Fax: +81 42 985 7209.

E-mail address: s_inoue@saitama-med.ac.jp (S. Inoue).

whether perfect HRE sequences function as hormone receptor binding sites *in vivo*, which regulate the transcription of steroid target genes.

Previously, we have identified perfect ARE sequences of IR3 type *in silico* in the human genome and characterized AR-binding ability of perfect AREs on chromosome X [3]. More than one-third of perfect ARE sequences (8 of 21) recruited more ARs compared with the proximal promoter region of prostate-specific antigen (PSA) containing a functional ARE. In the present study, we extend our study to a question whether perfect HRE sequences *in silico* situated in the proximal upstream regions of annotated genes function as bona fide functional binding sites for steroid hormone receptors in human cells derived from different tissues. Our combined approach of bioinformatics and experimental validation identified several perfect HRE sequences that could bind to steroid hormone receptors and some proximal downstream genes that were responsive to hormone treatment in the vicinity of the perfect HREs.

Materials and methods

Bioinformatics. Consensus HREs in the human genome (Ensembl version 34 based on the NCBI 35 assembly retrieved from Ensembl ftp site [4]) were screened by a 'pipeline' computational system called SayaMatcher [5], which utilizes in-house Perl script and a program for regular expression search of a nucleotide sequence (program name: dreg) in EMBOSS package [6]. The regular expression pattern for HRE was obtained from a recent literature by Nelson et al. [7], in which the palindromic 5'-AGAACA_nnnTGGTCT-3' sequence corresponding to the ARE sequence in TRANSFAC database [8] was used as a consensus sequence.

Cell culture. Human prostate cancer LNCaP cells and DU145 cells, breast cancer T47D cells, and osteosarcoma SaOS2 cells were purchased from American Type Culture Collection (Rockville, MD). LNCaP cells were maintained in RPMI 1640 supplemented with 4.5 g/dl glucose, 1 mM sodium pyruvate, 10 mM Hepes, and 10% fetal bovine serum (FBS). Other cells were maintained in DMEM supplemented with 10% fetal bovine serum (FBS). Prior to hormone addition, cells were cultured for 2 days in phenol red-free RPMI 1640 or DMEM supplemented with 5% dextran-charcoal stripped FBS (dcc-FBS) and 1 day in phenol red-free medium supplemented with 2.5% dcc-FBS.

Chromatin immunoprecipitation assay. Chromatin immunoprecipitation (ChIP) assay was performed as described previously [3]. LNCaP cells, T47D cells, and DU145 cells as well as SaOS2 cells after 72-h hormone depletion were treated with 10 nM R1881 (NEN Life Science Products, Boston, MA), progesterone (Sigma), or dexamethasone (Sigma), respectively, for indicated times. Control cells were treated with 0.1% ethanol as a vehicle. Cells were fixed in 1% formaldehyde for 5 min at room temperature. Chromatin was sheared to an average size of 500 bp by sonication using a Bioruptor ultrasonicator (Cosmo-Bio, Tokyo, Japan). Lysates corresponding to 2×10^7 cells were rotated at 4 °C for overnight with 3 μ g of polyclonal antibodies against AR (H-280, Santa Cruz Biotechnology, Santa Cruz, CA), PR (H-190, Santa Cruz Biotechnology), or GR (H-300, Santa Cruz Biotechnology). Salmon sperm DNA/protein A-agarose (Upstate Biotechnology, Lake Placid, NY) was added and incubated for 1 h. Washing and reversal of cross-links was performed as described [9]. Precipitated DNA fragments were quantified by quantitative real-time PCR using the Applied Biosystems 7000 sequence detector (Foster City, CA) based on SYBR Green I fluorescence. Primer pairs were designed by Primer Express ver. 2.0 software (Applied Biosystems), generating perfect ARE-containing fragments with the requirements of primer T_m temperature at basically 58–60 °C and the requirements of amplicon length for 50–150 bp. The protocol of PCR was 2 min at 50 °C, 10 min at 95 °C, and 40 cycles of

15 s at 95 °C and 1 min at 60 °C. To determine relative differences among the treatment groups for the ChIP assays, we used the $\Delta\Delta C_T$ method as outlined in the Applied Biosystems protocol for reverse transcriptase-PCR (<http://www.appliedbiosystems.com/>). The threshold cycle C_T was defined as the cycle at which the fluorescence signal was statistically significant over background. The reciprocal of 2^{C_T} (used C_T as an exponent for the base 2) for each target element was normalized by that for an external control of the genomic fragment GAPDH and subsequently normalized by input DNAs obtained from either steroid or vehicle-treated cells. Fold enrichment in steroid-treated cells was obtained through the division by the values in vehicle-treated cells. Genomic fragments containing proximal or distal ARE in the promoter region of prostate-specific antigen (PSA) [–250 to –39 bp and –4170 to –3978 bp from the transcriptional initiation site (TSS), respectively] [3,9] were used as positive controls for AR binding. The proximal promoter region of α subunit of amiloride-sensitive epithelial sodium channel (ENaC α) (–867 to –761 bp from the TSS, containing the GRE sequence as previously described [10]) was used as a positive control for GR binding. We also used a genomic fragment on the intron 1 of FK506 binding protein 51 (FKBP51) (+359 to +426 bp from the TSS of the long isoform ENST00000357266) as a positive control for GR and PR binding, which was newly identified in our ChIP assay scanning the gene regulatory region of FKBP51. The quality of precipitated DNA was analyzed by PCR amplification of positive controls and 8% polyacrylamide electrophoresis prior to quantitative analyses and the batches of ChIP samples with maximal responses were selected for quantitative PCR. The sequences of the primers used in ChIP assays (synthesized by Sigma Genosys, Japan) are described in Table 1.

Quantitative reverse transcription-PCR. Total RNA was extracted from hormone-treated or 0.1% ethanol-treated cells for indicated times using ISOGEN reagent (Nippon Gene, Tokyo, Japan). First strand cDNA was generated from RNase-free DNase I-treated total RNA by using SuperScript II Reverse Transcriptase (Invitrogen, Carlsbad, CA) and pdT_{12–18} primer (Amersham Biosciences, Piscataway, NJ). Hormone responsiveness of the proximal genes located downstream to the 26 perfect HRE sequences was analyzed by quantitative reverse transcription-PCR (RT-PCR) using the Applied Biosystems 7000 sequence detector based on SYBR Green I fluorescence. Primer design and PCR protocol were as described above. The evaluation of relative differences of PCR product amounts among the treatment groups was carried out using the $\Delta\Delta C_T$ method. The reciprocal of 2^{C_T} (used C_T as an exponent for the base 2) for each target gene was normalized by that for GAPDH coding region, followed by the comparison with the relative value in vehicle-treated cells. PSA was served as a positive control for androgen responsiveness, while FKBP51 was used for both progesterone and glucocorticoid responsiveness. ENaC α was another positive control for the glucocorticoid response and the progesterone-dependent change of ENaC α mRNA level was also evaluated. The primer sequences for the amplifications are described in Table 2.

Results

In silico identification of perfect palindromic HRE sequences in the human genome

In terms of palindromic HRE sequences composed of two AGACA sequences separated by a 3-bp spacer, only a few near-consensus sequences have been identified in the vicinity of human genes that are responsive to steroid hormones including androgen, progesterone, glucocorticoid, and mineralocorticoid. In order to answer the question of whether perfect palindromic HRE sequences do function as *in vivo* binding sites for steroid hormone receptors, we computationally searched for all the consensus HRE sequences in the human genome utilizing in-house Perl

Table 1
Primers for quantitative ChIP assay

Target	Forward primer	Reverse primer
HRE sites		
HRE1	5'-CCAAATATGTCCATTTCATCCAACA-3'	5'-GGAAACATACGCATTGCCTAGAA-3'
HRE2	5'-CCAACAAACACTAGCAGAACATTATG-3'	5'-AGAGGGCATCTGAACAATAGAAGA-3'
HRE3	5'-CAAGGCTGGGGACATTTAA-3'	5'-CGAAGCTGCTGGGAGGTATC-3'
HRE4	5'-TCAAAGGTGGTGCTTATCAT-3'	5'-GAACTGAGTTCCAGGAAATATG-3'
HRE5	5'-AAGGGAAAAGATCTCATATTGCA-3'	5'-CACGGCAATTTTAATACTCATCA-3'
HRE6	5'-GACCATACCTTGGTATGACCCCTT-3'	5'-TCTGTGTTTATCTGCAAGTCAGT-3'
HRE7	5'-AAGAAACCATTTTGAATGTGCC-3'	5'-CAGGTGAGTTAGGCTCTGATAAAACC-3'
HRE8	5'-TACTTACCCAGTGCTTCCCAATC-3'	5'-GAGCACCTGAGGTGTTTTTTTTT-3'
HRE9	5'-AGCGAGACTCCGTCTTAAA-3'	5'-GGAGGAGTTACTGTTACAATTAT-3'
HRE10	5'-ACACTGTGGTGGGCAGGATT-3'	5'-GCTCAACAAACAGGCTAAGACA-3'
HRE11	5'-TGTCTTCTGAAGACTTCCATTC-3'	5'-TACAGGCACACACCACCAGATTG-3'
HRE12	5'-CTTGCACTGAGCCGAGATC-3'	5'-TTCTGGAGGAGTTACTGTTACAATTA-3'
HRE13	5'-CCTGCTCTTTGATTTTCGTAAGA-3'	5'-TGAAGCCAAATGCTGAAGTG-3'
HRE14	5'-TTGACAGTATATTTAGAGCGTGTATCT-3'	5'-GTAAATTCTCTTCCCTATGGTAGAAT-3'
HRE15	5'-TGTGCAAGAAATGCCATCCT-3'	5'-CATCATAGCTGCATGGGTGTT-3'
HRE16	5'-TGAAAATAATTTCACTTATCCTTTAAGC-3'	5'-AATGACTTGCTGACCAATCC-3'
HRE17	5'-GGCCTATTTTAAATCAGCATTTTCA-3'	5'-ATCCTCATTCCTACCCAGCATCT-3'
HRE18	5'-ACATCAGTGGCTCTTAAGCACTG-3'	5'-CAGAGAGAATAAAAAACTAGCCCAA-3'
HRE19	5'-AATGCTGCTCAAAGAAATCAGA-3'	5'-TTTAATTCCAAATGCATTTTAGAAC-3'
HRE20	5'-CTCCATGTTTGAGGCGGAGA-3'	5'-CATCCCACTTTGTGCGTAACCTT-3'
HRE21	5'-CTCCAGGGCTGTACTGGTATCTG-3'	5'-CAGGTGTGGGCCCTCTGT-3'
HRE22	5'-TCTTAAGCAAAAGGCCAACA-3'	5'-TCTATCATTACAAGAAAAGTTGGTTATCTG-3'
HRE23	5'-GCCAGCACTCACGTGGCTAT-3'	5'-CTGCGGTGCCGTGACA-3'
HRE24	5'-TGTAACACAAAACTGCTAGAACACTCT-3'	5'-GCTCAGCCATGGGTGGT-3'
HRE25	5'-GGGAGCTGGCTAGAACACTCA-3'	5'-CATGCTGGGCTAAAGCAGAAC-3'
HRE26	5'-TGTTGTGCTCAAAGAGTAGATTGG-3'	5'-TCATCACTTATTGCTGAACTATCA-3'
FKBP51 intron 1	5'-TGCAAGAGCGGTTGATCTG-3'	5'-GCGAGAACAGCCTGACTCA-3'
ENaC α promoter	5'-CCTCTGGTTGCCACATTC-3'	5'-GGGCCCTAGGACATTCTGTT-3'
PSA proximal promoter	5'-TCTGCCTTTGTCCCTAGAT-3'	5'-AACCTTCATTCCCAGGACT-3'
PSA distal promoter	5'-ACAGACCTACTCTGGAGGAAC-3'	5'-AAGACAGCAACACCTTTTT-3'

script and a program for regular expression pattern search of a nucleotide sequence in the EMBOSS package (program name: dreg) as previously described [3,5]. The screening defined 563 sequences on the NCBI 34 assembly of the human genome [3] and additional two sites on the NCBI 35 assembly. The two new sequences were located on chromosomes 1 and 14. The number of HRE sites was larger than the expected frequency in random DNA sequences as calculated by the total number of base pairs in the genome divided by the frequency of a sequence with specific base pairs at 12 positions ($3272 \text{ Mb}/4^{12} = 195$). The distribution of consensus sequences among the chromosomes is generally consistent with chromosomal size [3]. The highest frequency of HRE was 28.0 sites per 100 Mb on chromosome 8 and the lowest frequency was 6.9 sites per 100 Mb on chromosome Y, the average being 17.4 ± 4.4 sites per 100 Mb.

Perfect HREs located in the proximal upstream regions of annotated genes

To investigate whether the computationally identified perfect HRE sequences located in the canonical proximal promoters of annotated genes particularly function as steroid hormone receptor binding sites, we selected perfect HREs lying within 10 kb upstream to the transcriptional

start sites (TSSs) of known human genes on the Ensembl Genome Browser [4]. We obtained 26 of 565 perfect HRE sequences (4.6%) in the proximal promoter regions of annotated genes (Table 3). The distribution of 26 perfect HREs in the proximal promoter elements was limited to 14 of 24 chromosomes and rather independent of chromosomal size. The longest chromosome 1 contained 5 sequences, while 4 sequences were found in both chromosomes 6 and 11.

In vivo recruitment of steroid hormone receptors to perfect HREs located 5' to annotated genes

To assess whether these perfect HREs situated within 10 kb of the 5' ends of known genes (5' HREs) do recruit steroid hormone receptors in vivo, we performed ChIP assay in human cells endogenously expressing receptors. Hormone-depleted cells were treated for 24 h (also 2 h in T47D cells) with either 10 nM of steroids or 0.1% ethanol as a vehicle, cross-linked by formaldehyde, sonicated, and immunoprecipitated by specific polyclonal antibodies against steroid receptors. LNCaP cells were treated with R1881 (or vehicle) and immunoprecipitated by AR antibody; T47D cells were treated with progesterone and immunoprecipitated by PR antibody; and DU145 cells as well as SaOS cells were treated with dexamethasone and

Table 2
Primers for quantitative RT-PCR

Target	Adjacent HRE	Forward primer	Reverse primer
IQSEC2	HRE1	5'-ACATCACAGAACTTGAGGACTCCTT-3'	5'-AGGGCTTCGTCGATGGATTC-3'
MELL1	HRE2	5'-GAGATCACGCACGGCTTTG-3'	5'-CCAATCCATCATGTTGCCATT-3'
HS163_HUMAN	HRE3	5'-GATGGAGGCGGTGGTGTTC-3'	5'-GTAATTATGAAGTAGACCGAGAGGAAGAT-3'
DISP1	HRE4	5'-CATGGAGCTGGAAGGAGTACAG-3'	5'-CACCAGCTGCCCTGATAGTATCATT-3'
EXTL2	HRE5	5'-AGGCAGCCACGTGTTCTATTG-3'	5'-TTAACGCAGCAAAACGATCACT-3'
NP_689710.1	HRE6	5'-GAGGGCAGTGGCGTGATC-3'	5'-CAGTAGAATGGCGTGAAGTCAAG-3'
B3GALT3	HRE7	5'-TGCAGCCCCATGGCTTTTC-3'	5'-GGCATGTGGTGTTCCTTAGCA-3'
SLC36A1	HRE8	5'-CGGAAGGCCCTCAACAACCT-3'	5'-CCATGTTGTGCTATTGCTTTGAC-3'
DAXX	HRE9	5'-GCCCTTCACCAGTGTCTTAGAGA-3'	5'-GAGACGCCTCCATTGAAGGA-3'
ULBP3	HRE10	5'-CGGACTGACCACCTTCTTCAA-3'	5'-TGCATCAGGAAGTCCCTAAGC-3'
KIF13A	HRE11	5'-GCTGCTATGGAGTTAGTATTACGAAAAC-3'	5'-CTTCAAACCTGCGTGAAACTCTGT-3'
ZNF297	HRE12	5'-CCTCTCAAGGATGCGCTCTT-3'	5'-GGGAATTTCCACGGGAAGTC-3'
GIMAP5	HRE13	5'-CAGGTGAAAACAGGAACATGGA-3'	5'-GGCCTGTGACTCAAGATGGA-3'
ZNF398	HRE14	5'-GAGCTTCCGCTACAAACAGACA-3'	5'-ACCCACAGCCCTCCATT-3'
NP_055096.2	HRE15	5'-CAGTGACCTCTGCGTGACA-3'	5'-AACCCAGAGGAAACAAGGAACA-3'
Q9C0D7_HUMAN	HRE16	5'-CCAGCTGTCTGCGGACTAC-3'	5'-AGTCGAGAGTCAGAGATGCTGTCTATT-3'
OR8D4	HRE17	5'-GCTTTTATCCTCACCAGCATCCT-3'	5'-GAGCTACAGGTGCTAAACGCTTT
ENSG00000182203	HRE18	5'-ACACCAACAGGATACACTGAAAGC-3'	5'-GCACATTTAACTTGCAGGTTTTTG-3'
ENSG00000185439	HRE19	5'-TGCCTTCATGGTGGACAATG-3'	5'-GGATGCTCGATGTCCAGGTT-3'
CV106_HUMAN	HRE20	5'-CACAATTTAACTTACGAACCTCCAAC-3'	5'-GCCAAGTGAAGCTCTAGTTAATGTCTTG-3'
SRP14	HRE21	5'-GTTGTTGGAGAGCGAGCAGTT-3'	5'-CCGACGTCCGGCACTTC-3'
NP_874362.2	HRE22	5'-TGCAGCGACTTGTGGACATC-3'	5'-CGAATGCAGGGCAGTCAGA-3'
MVD	HRE23	5'-TGGCGGCAGTCACCTGTACA-3'	5'-CGCGCTTGCCCCAGTA-3'
ENSG00000189289	HRE24	5'-GCCTGCTCCACCCAGAGAA-3'	5'-AACAGGTTCCAGCAGCTCAGA-3'
NP_775751.1	HRE25	5'-CACCAAGTGCAGGACAGTCTTC-3'	5'-GGTGCTTTTCGTCCACAGTGA-3'
NP_689725.2	HRE26	5'-CTTTACTAGCAACACCTCCTTCCAT-3'	5'-TGCACCCGAAAGATTACAAACTT-3'
FKBP51 coding		5'-CTGCAGAGATGTGGCATTCACT-3'	5'-TCCAGAGCTTTGTCAATTCCAA-3'
ENaCα coding		5'-CGCATGAAGACGGCCTTCT-3'	5'-CGCATGAAGACGGCCTTCT-3'
PSA coding		5'-GCCCTGCCCCGAAAGG-3'	5'-GATCCACTTCCGTAATGCA-3'
GAPDH coding		5'-GGTGGTCTCTCTGACTTCAACA-3'	5'-GTGGTCTGTGAGGCAATG-3'

Table 3
Perfect HRE sites that locate upstream to annotated genes at a distance within 10 kb

HRE ID	Chr. location	Start position	Stop position	HRE sequence	Proximal gene	Ensembl ID	Distance (kb)
HRE1	Xp11.22	53237415	53237429	AGAACAAttaGTTTCT	IQSEC2	ENSG00000124313	3.9
HRE2	1p36.32	2603985	2603999	AGAACAagcTGTTC	MELL1	ENSG00000142606	7.4
HRE3	1q42.12	220841360	220841374	AGAACAgttGTTTCT	HS163_HUMAN	ENSG00000143771	9.9
HRE4	1q41	219495172	219495186	AGAACAgtatGTTTCT	DISP1	ENSG00000154309	9.4
HRE5	1p21.1	101079768	101079782	AGAACAtaaGTTTCT	EXTL2	ENSG00000162694	6.2
HRE6	1p36.11	25913162	25913176	AGAACAgtcTGTTC	NP_689710.1	ENSG00000177493	9.9
HRE7	3q26.1	162310806	162310820	AGAACAAttcGTTTCT	B3GALT3	ENSG00000169255	5.0
HRE8	5q33.1	150798127	150798141	AGAACAagcgTGTTC	SLC36A1	ENSG00000123643	9.2
HRE9	6p21.32	33401650	33401664	AGAACAAttcTGTTC	DAXX	ENSG00000007565	3.0
HRE10	6q25.1	150487454	150487468	AGAACAataGTTTCT	ULBP3	ENSG00000131019	5.1
HRE11	6p22.3	18101058	18101072	AGAACAAtgcTGTTC	KIF13A	ENSG00000137177	5.4
HRE12	6p21.32	33401650	33401664	AGAACAAttcTGTTC	ZNF297	ENSG00000168351	8.2
HRE13	7q36.1	149868965	149868979	AGAACAcaatGTTTCT	GIMAP5	ENSG00000196329	3.2
HRE14	7q36.1	148274989	148275003	AGAACAAttaGTTTCT	ZNF398	ENSG00000197024	7.5
HRE15	8q24.3	144987498	144987512	AGAACAaccgTGTTC	NP_055096.2	ENSG00000179950	8.9
HRE16	11q22.3	109509863	109509877	AGAACAAtttGTTTCT	Q9C0D7_HUMAN	ENSG00000149289	1.9
HRE17	11q24.1	123276891	123276905	AGAACAaagTGTTC	OR8D4	ENSG00000181518	5.5
HRE18	11q24.2	125433385	125433399	AGAACAaaaTGTTC	ENSG00000182203	ENSG00000182203	9.3
HRE19	11q12.3	62569237	62569251	AGAACAaatGTTTCT	ENSG00000185439	ENSG00000185439	1.8
HRE20	14q21.2	44792647	44792661	AGAACAAtttGTTTCT	CV106_HUMAN	ENSG00000129534	0.5
HRE21	15q15.1	38128202	38128216	AGAACAaaatGTTTCT	SRP14	ENSG00000140319	9.5
HRE22	15q22.3	62987265	62987279	AGAACAgaatGTTTCT	NP_874362.2	ENSG00000166839	3.9
HRE23	16q24.3	87266241	87266255	AGAACAactGTTTCT	MVD	ENSG00000167508	9.2
HRE24	17q21.32	44836359	44836373	AGAACAActcTGTTC	ENSG00000189289	ENSG00000189289	3.2
HRE25	19p13.3	2844818	2844832	AGAACAaggaTGTTC	NP_775751.1	ENSG00000171970	7.1
HRE26	22q13.1	38616692	38616706	AGAACAgaatGTTTCT	NP_689725.2	ENSG00000176177	2.4

immunoprecipitated by GR antibody. We performed quantitative PCR using the genomic DNA obtained from each cell line (Fig. 1). The proximal and distal promoter regions of PSA including ARE sequences were used as positive controls for the AR association [3], and the proximal promoter of ENaC α including HRE sequences [10] was used as a positive control for the GR. The proximal promoter fragment of ENaC α was also found to recruit PR in the present study, as the PR recruitment to the fragment was \sim 9.5- and \sim 20-fold versus vehicle by 2 and 24-h treatments, respectively (Fig. 1B). FKBP51 has been previously described as a glucocorticoid or progesterone-inducible gene [11,12], however, functional HREs in the gene regulatory region have not been well characterized. We newly identified a functional HRE in the intron 1 of FKBP51 through scanning the gene regulatory region by ChIP assay. The intronic HRE of FKBP51 recruited PR in T47D cells as well as GR in DU145 cells and SaOS2 cells (Figs. 1B–D).

Of 26 genomic fragments containing perfect 5' HRE sequences, 14 fragments recruited either one of the three steroid receptors by >2 -fold enrichment in steroid-treated cells versus vehicle-treated cells (Fig. 1). HRE20 recruited three different receptors while HRE1, HRE15, and HRE24 recruited two different receptors by >2 -fold enrichment. Other HRE sequences preferentially bound to only one steroid receptor. There seems to be no particular specificity of 3-bp spacer sequences for the receptor recruitment among the 14 functional HRE sequences.

Potential transcriptional regulation of proximal genes located downstream to perfect HREs

To examine whether the functional perfect 5' HREs regulate transcriptional activity of their downstream proximal genes, we next performed quantitative RT-PCR for the downstream genes in the vicinity of the functional perfect 5' HRE sequences. Of 7 proximal genes in the vicinity of functional perfect 5' HRE sequences that recruited AR by >2 -fold enrichment in LNCaP cells, OR8D4 (olfactory receptor family 8, subfamily D, member 4) adjacent to HRE17 exhibited a >2 -fold increase in the mRNA level androgen-dependently (Fig. 2A). Of nine proximal genes in the vicinity of functional perfect 5' HREs in T47D cells, MELL1 (mel transforming oncogene-like 1) and NP_775751.1 adjacent to HRE2 and HRE25, respectively, were upregulated by >2 -fold at the mRNA level by progesterone treatment (Fig. 2B). Neither gene adjacent to the 2 functional perfect 5' HREs in DU145 cells exhibited >2 -fold increase in mRNA level, whereas Q9C0D7 adjacent to HRE16 in SaOS2 cells was upregulated by >2 -fold at mRNA level glucocorticoid-dependently (Figs. 2C and D).

Discussion

A genome-wide in silico screening of palindromic perfect HREs identified 565 exact consensus sequences in the NCBI 35 assembly of the human genome. Of 565 perfect

HRE sequences, 26 sites were situated within 10 kb upstream to the TSS of annotated genes. We investigated whether these perfect 5' HRE sequences functioned as bona fide binding sites for steroid hormone receptors in different cell lines, and demonstrated that 14 of 26 perfect HREs significantly (>2 -fold) recruited some of the receptors by performing ChIP assay for AR, GR, and PR binding. Hormone-dependent upregulation of mRNA level by >2 -fold was shown in the four proximal genes adjacent to the functional perfect 5' HREs. The average distance between 5' HREs and the TSS of their proximal genes was not significantly smaller among these four proximal genes compared with the all 26 proximal genes in the vicinity of perfect 5' HREs; the former and latter values were 4.5 ± 2.4 and 6.0 ± 3.0 kb, respectively.

The four hormone-responsive proximal genes adjacent to functional 5' HRE sequences were upregulated by only either one of the three steroid hormone receptors and not by two or three other receptors. OR8D4 (HRE17) was identified as an AR target gene in LNCaP cells, MELL1 (HRE2) and NP_775751.1 (HRE25) were identified as PR target genes in T47D cells, and Q9C0D7 (HRE 16) was identified as a GR target gene in SaOS2 cells. Although a single IR3 type consensus HRE may be a potential binding site for various steroid hormone receptors, no common target gene for AR, PR, and GR was identified among the proximal genes close to the functional perfect 5' HREs by the functional criteria used in this study. It is also notable that no significant GR target gene was shown in DU145 cells up to 24-h dexamethasone treatment, though the mRNA level of NP_055096.2 (HRE15) was increased >2 -fold by 48-h treatment (data not shown). The results of DU145 and SaOS2 cells suggest that target genes for even a single steroid hormone receptor may be varied in different cell systems. Future studies including such as the chromosomal accessibility, the receptor recognition mechanism, and the coactivator recruitment in the surroundings of consensus HREs may reveal the specificity and the similarity of the transcription factor binding sites.

Besides the functions of the above four perfect HRE sequences adjacent to the steroid target genes, other near-consensus HRE sequences may participate in the regulation of gene expression. Using the SayaMatcher system [5] that enables us to visualize the genomic position for HREs on the Ensembl browser, we found several other near-consensus HRE sequences ($>80\%$ relative profile score threshold) based on the position-specific scoring matrix for ARE (MA0007) on the Jaspar database (<http://jaspar.cgb.ki.se/>), which is a collection of transcription factor DNA-binding preferences [13]. For instance, there is a near-consensus HRE at \sim 8-kb downstream to the 3'-end of MELL1. In the vicinity of Q9C0D7, one near-consensus HRE is in the intron 1 and the other is situated at 40-kb downstream to the 3'-end. OR8D4 contains a near-consensus HRE in its only exon. Although no particular high score position for the ARE matrix is found in the vicinity of NP_775751.1, there are a number of nuclear

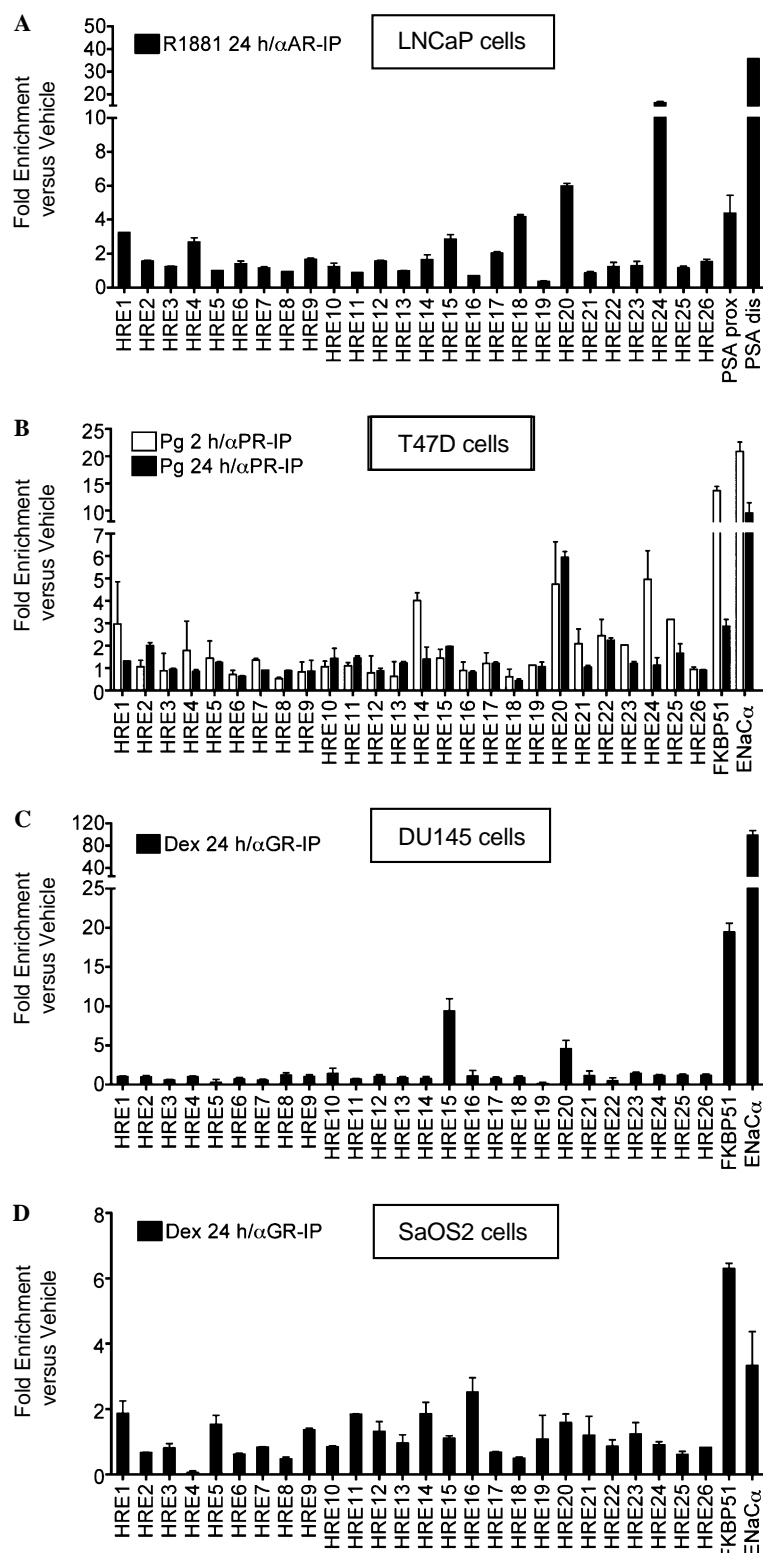


Fig. 1. In vivo recruitment of steroid hormone receptors to perfect HREs in the proximal upstream region of annotated genes. LNCaP, T47D, DU145, and SaOS2 cells after 72-h hormone depletion were treated with indicated hormones (10 nM each) or vehicle (0.1% EtOH) for indicated times. Immunoprecipitated DNA fragments were quantified by quantitative real-time PCR. In each case, fold enrichment values in hormone-treated samples precipitated by specific antibodies against receptors were normalized by those in vehicle-treated samples precipitated by the identical antibodies. Each result is the mean \pm SD of two independent experiments in duplication (four determinants). Prostate-specific antigen (PSA) proximal (prox) and distal (dis) promoter regions including HREs were used as positive controls for AR binding. FK506-binding protein, 51-kDa (FKBP51) intron 1, and epithelial sodium channel α subunit (ENaC α) proximal promoter regions were used as positive controls for PR and GR binding. (A) AR recruitment in prostate cancer LNCaP cells. (B) PR recruitment in breast cancer T47D cells. (C) GR recruitment in prostate cancer DU145 cells. (D) GR recruitment in osteosarcoma SaOS2 cells.

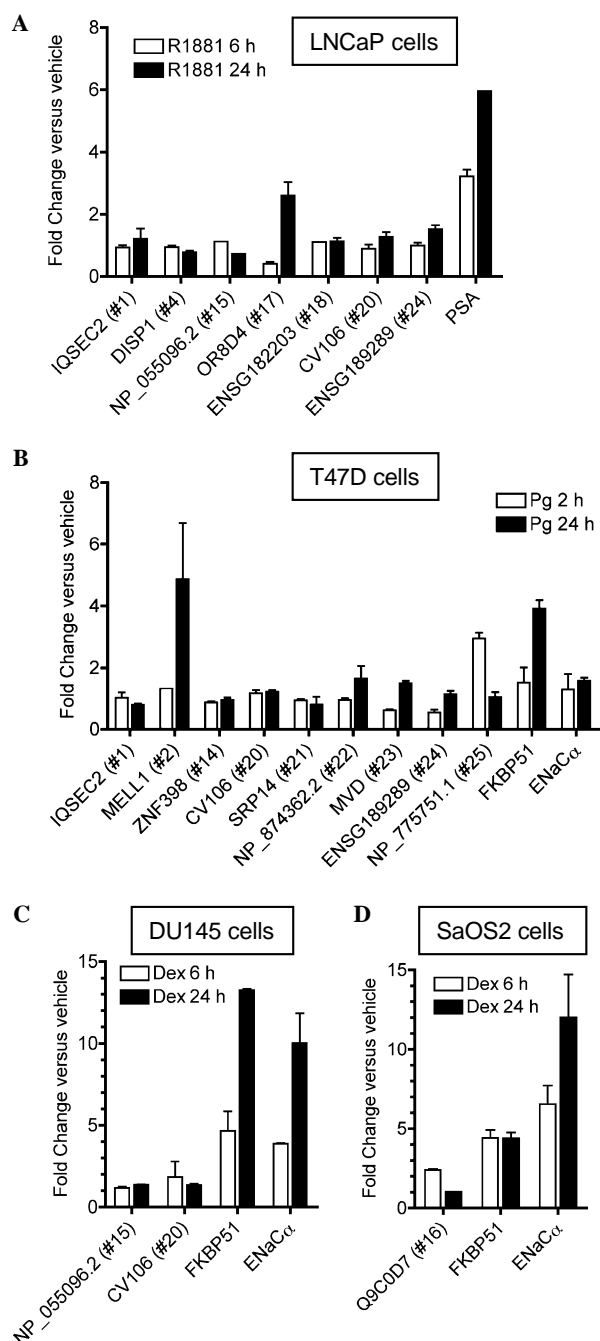


Fig. 2. Quantitative RT-PCR of proximal genes downstream to perfect HREs. LNCaP, T47D, DU145, and SaOS2 cells after 72-h hormone depletion were treated with indicated hormones (10 nM each) or vehicle (0.1% EtOH) for indicated times. Real-time PCR was conducted using the first strand cDNAs generated from the total RNAs of these cells. Each result is the mean \pm SD of two independent experiments in duplication (four determinants). PSA was served as a positive control for androgen responsiveness. FKBP51 and ENaC α were served as positive controls for progesterone and glucocorticoid responsiveness. The numbers in parentheses after annotations correspond to the HRE ID in the vicinity of target genes. (A) R1881-dependent gene expression in LNCaP cells. (B) Progesterone-dependent gene expression in T47D cells. (C) Dexamethasone-dependent gene expression in DU145 cells. (D) Dexamethasone-dependent gene expression in SaOS2 cells.

receptor binding sites in the surroundings of the gene, suggesting that the region may be active for transcriptional regulation. Future studies will reveal whether these near-consensus HRE sequences coordinately function with the perfect HRE in the transcriptional regulation of the proximal genes.

Regarding the new steroid target genes identified in the present study, none of the genes has been characterized in connection with steroid hormones. The AR target gene OR8D4 (UniGene Hs. 449688) encodes a putative G protein-coupled receptor that belongs to the olfactory receptor superfamily, which is the largest gene family in the mammal genomes. In the mouse genomes, \sim 1300 genes were identified, whereas the number of olfactory receptors is expected \sim 500–750 genes in human [14]. Although ligands for most of the olfactory receptors have not been identified, it is plausible that certain receptors for odors or pheromones may be regulated by sex steroid hormones including androgen. The PR target gene MELL1 (UniGene Hs. 546429) is a membrane-bound zinc metalloprotease. It has been recently shown that mouse ortholog NL1 is expressed mainly in the testis as a secreted protein and male mice deficient in the NL1 exhibited reduced fertility [15]. It seems to be an interesting question whether human MELL1 may also play a role in reproductive tissues, particularly in the organs regulated by the progesterone-mediated gene network.

Another PR target gene NP_775751.1 encodes a putative KRAB domain-containing zinc finger protein. KRAB-containing zinc finger proteins are characterized to participate in the regulation of cell differentiation and development through transcriptional repression of RNA polymerase promoters, RNA binding, and RNA splicing [16]. Although many of these proteins have been identified, little is known about their structure and function. Since NP_775751.1 is clustered with a number of KRAB-containing zinc finger proteins on 19p13.3, these KRAB-containing zinc finger proteins might have arisen from a divergence event during evolution and may function coordinately in the various developmental and differentiation stages of organs. The GR target gene Q9C0D7 in SaOS2 cells is also annotated as zinc finger CCCH-type containing 12C (UniGene Hs. 376289). The CCCH-type zinc finger proteins such as tristetraprolin has been known as a binding protein to AU-rich element containing RNA transcripts [17], though no particular function of Q9C0D7 has yet been characterized.

In the present study, we focused on the perfect HRE sequences in the proximal upstream regions of annotated genes and found some of the elements were functional binding sites for steroid hormone receptors. It has been recently revealed, however, that transcription factor binding sites in intronic regions or in the 3' downstream regions are also important for transcriptional regulation. Indeed, half of the perfect intronic ARE sequences on chromosome X (5 of 10) recruited more ARs compared with the proximal ARE on the PSA promoter in our

previous study [3]. In this study, we found that the FKBP51 intronic HRE sequence exhibited a significant binding ability for GR and PR. In the case of transcription factor binding sites on chromosomes 21 and 22, 36% of these regions are situated within known genes or proximal to the 3' most exon of a gene and the frequency was also higher than that of binding sites within 5' to known genes (22%) [18]. Thus, the further characterization of HRE sequences in introns as well as 3' downstream regions will be required to reveal the whole entity of the transcriptional regulation mechanism mediated through steroid hormone receptors.

In summary, we have presented an integrated strategy for exploring novel steroid target genes that possess a functional perfect HRE sequence in their proximal upstream regions. Our results have shown that more than half of the perfect HREs identified by in silico analysis were functional binding sites for some of the steroid receptors including AR, PR, and GR. Potential steroid target genes were identified in the vicinity of perfect HRE sequences. We expect our strategy will be useful for further studies of gene regulation and the elucidation of the steroid hormone receptor-mediated gene network that exerts distinct physiological and pathophysiological functions.

Acknowledgments

We thank Y. Suzuki and R. Nozawa for their technical assistance. This work was supported in part by Grants-in-Aid from the Ministry of Health, Labor and Welfare; from the Japan Society for the Promotion of Science. This work was supported in part by a grant of the Genome Network Project from the Ministry of Education, Culture, Sports, Science and Technology of Japan and for Development of New Technology from The Promotion and Mutual Aid Corporation for Private Schools of Japan.

References

- [1] D.J. Mangelsdorf, C. Thummel, M. Beato, P. Herrlich, G. Schutz, K. Umesono, B. Blumberg, P. Kastner, M. Mark, P. Chambon, R.M. Evans, The nuclear receptor superfamily: the second decade, *Cell* 83 (1995) 835–839.
- [2] M. Beato, P. Herrlich, G. Schutz, Steroid hormone receptors: many actors in search of a plot, *Cell* 83 (1995) 851–857.
- [3] K. Horie-Inoue, H. Bono, Y. Okazaki, S. Inoue, Identification and functional analysis of consensus androgen response elements in human prostate cancer cells, *Biochem. Biophys. Res. Commun.* 325 (2004) 1312–1317.
- [4] T. Hubbard, D. Andrews, M. Caccamo, G. Cameron, Y. Chen, M. Clamp, L. Clarke, G. Coates, T. Cox, F. Cunningham, et al., Ensembl 2005, *Nucleic Acids Res.* 33 (2005) D447–D453.
- [5] H.U. Bono, SayMatcher: Genome scale organization and systematic analysis of nuclear receptor response elements, *Gene* (2005). Available online August 24.
- [6] P. Rice, I. Longden, A. Bleasby, EMBOSS: the European Molecular Biology Open Software Suite, *Trends Genet.* 16 (2000) 276–277.
- [7] P.S. Nelson, N. Clegg, H. Arnold, C. Ferguson, M. Bonham, J. White, L. Hood, B. Lin, The program of androgen-responsive genes in neoplastic prostate epithelium, *Proc. Natl. Acad. Sci. USA* 99 (2002) 11890–11895.
- [8] V. Matys, E. Fricke, R. Geffers, E. Gossling, M. Haubrock, R. Hehl, K. Hornischer, D. Karas, A.E. Kel, O.V. Kel-Margoulis, D.U. Kloos, S. Land, B. Lewicki-Potapov, H. Michael, R. Munch, I. Reuter, S. Rotert, H. Saxel, M. Scheer, S. Thiele, E. Wingender, TRANSFAC: transcriptional regulation, from patterns to profiles, *Nucleic Acids Res.* 31 (2003) 374–378.
- [9] Y. Shang, M. Myers, M. Brown, Formation of the androgen receptor transcription complex, *Mol. Cell* 9 (2002) 601–610.
- [10] V.E. Mick, O.A. Itani, R.W. Loftus, R.F. Husted, T.J. Schmidt, C.P. Thomas, The alpha-subunit of the epithelial sodium channel is an aldosterone-induced transcript in mammalian collecting ducts, and this transcriptional response is mediated via distinct cis-elements in the 5'-flanking region of the gene, *Mol. Endocrinol.* 15 (2001) 575–588.
- [11] H. Vermeer, B.I. Hendriks-Stegeman, B. van der Burg, S.C. van Buul-Offers, M. Jansen, Glucocorticoid-induced increase in lymphocytic FKBP51 messenger ribonucleic acid expression: a potential marker for glucocorticoid sensitivity, potency, and bioavailability, *J. Clin. Endocrinol. Metab.* 88 (2003) 277–284.
- [12] T.R. Hubler, W.B. Denny, D.L. Valentine, J. Cheung-Flynn, D.F. Smith, J.G. Scammell, The FK506-binding immunophilin FKBP51 is transcriptionally regulated by progestin and attenuates progestin responsiveness, *Endocrinology* 144 (2003) 2380–2387.
- [13] A. Sandelin, W. Alkema, P. Engstrom, W.W. Wasserman, B. Lenhard, JASPAR: an open-access database for eukaryotic transcription factor binding profiles, *Nucleic Acids Res.* 32 (2004) D91–D94.
- [14] X. Zhang, S. Firestein, The olfactory receptor gene superfamily of the mouse, *Nat. Neurosci.* 5 (2002) 124–133.
- [15] M. Carpentier, C. Guillemette, J.L. Bailey, G. Boileau, L. Jeannotte, L. DesGroseillers, J. Charron, Reduced fertility in male mice deficient in the zinc metalloproteinase NL1, *Mol. Cell. Biol.* 24 (2004) 4428–4437.
- [16] R. Urrutia, KRAB-containing zinc-finger repressor proteins, *Genome Biol.* 4 (2003) 231.1–231.8.
- [17] E. Carballo, W.S. Lai, P.J. Blackshear, Feedback inhibition of macrophage tumor necrosis factor- α production by tristetraprolin, *Science* 281 (1998) 1001–1005.
- [18] S. Cawley, S. Bekiranov, H.H. Ng, P. Kapranov, E.A. Sekinger, D. Kampa, A. Piccolboni, V. Sementchenko, J. Cheng, A.J. Williams, R. Wheeler, B. Wong, J. Drenkow, M. Yamanaka, S. Patel, S. Brubaker, H. Tammana, G. Helt, K. Struhl, T.R. Gingeras, Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs, *Cell* 116 (2004) 499–509.