Cell- and gene-specific regulation of primary target genes by the androgen receptor

Eric C. Bolton,1 Alex Y. So,1,5 Christina Chaivorapol,2,4,6 Christopher M. Haqq,3 Hao Li,2,4,6 and Keith R. Yamamoto1,5,7

1Department of Cellular and Molecular Pharmacology, University of California, San Francisco, California 94143, USA; 2Department of Biochemistry and Biophysics, University of California, San Francisco, California 94143, USA; 3Department of Urology, University of California, San Francisco, California 94143, USA; 4California Institute for Quantitative Biomedical Research, University of California, San Francisco, California 94143, USA; 5Chemistry and Chemical Biology Graduate Program, University of California, San Francisco, California 94143, USA; 6Graduate Program in Biological and Medical Informatics, University of California, San Francisco, California 94143, USA

The androgen receptor (AR) mediates the physiologic and pathophysiologic effects of androgens including sexual differentiation, prostate development, and cancer progression by binding to genomic androgen response elements (AREs), which influence transcription of AR target genes. The composition and context of AREs differ between genes, thus enabling AR to confer multiple regulatory functions within a single nucleus. We used expression profiling of an immortalized human prostate epithelial cell line to identify 205 androgen-responsive genes (ARGs), most of them novel. In addition, we performed chromatin immunoprecipitation to identify 524 AR binding regions and validated in reporter assays the ARE activities of several such regions. Interestingly, 67% of our AREs resided within ~50 kb of the transcription start sites of 84% of our ARGs. Indeed, most ARGs were associated with two or more AREs, and ARGs were sometimes themselves linked in gene clusters containing up to 13 AREs and 12 ARGs. AREs appeared typically to be composite elements, containing AR binding sequences adjacent to binding motifs for other transcriptional regulators. Functionally, ARGs were commonly involved in prostate cell proliferation, communication, differentiation, and possibly cancer progression. Our results provide new insights into cell- and gene-specific mechanisms of transcriptional regulation of androgen-responsive gene networks.

[Keywords: Androgen receptor (AR); androgen response element (ARE); transcription; steroid receptor; chromatin immunoprecipitation (ChIP); prostate cancer]

Supplemental material is available at http://www.genesdev.org.

Received April 23, 2007; revised version accepted July 6, 2007.

Androgens are steroid hormones that control male development and reproductive function and are key regulators of prostate cancer cell growth and proliferation. They bind to the androgen receptor [AR], a member of the intracellular receptor family of transcriptional regulators [Chang et al. 1988; Lubahn et al. 1988], which in turn associates with genomic androgen response elements [AREs], in particular, AR can specifically recognize a 15-bp partially palindromic AR binding sequence (ARB5) motif [Claessens et al. 2001]. However, localization of AR to specific genomic sites is clearly more complex than simple sequence recognition. Similarly, AR can regulate transcription by distinct mechanisms; depending on cell or gene context, AR can activate or repress transcription, can functionally interact with other transcriptional regulators [Murtha et al. 1997; Sato et al. 1997] or with other intracellular receptors [Chen et al. 1997; Lee et al. 1999; Panet-Raymond et al. 2000; Chuang et al. 2005]. Hormone response element (HRE) organization is likely a major determinant of receptor specificity and coregulator recruitment in a given chromosomal context [Lefstin and Yamamoto 1998]. Thus, a detailed understanding of androgen action will require identification of genes bound and regulated by AR, identification of coregulators that interact with AR and characterization of the loci where these interactions occur.

RNA expression profiling of LNCaP prostate cancer cells has described ~500 transcripts whose levels are altered by androgen treatment [DePrimo et al. 2002; Nelson et al. 2002]. However, the mechanisms of regulation are nearly unexplored because the primary androgen responsive genes [ARGs], i.e., the subset regulated directly
by AR-occupied AREs, have not been defined. The primary ARGs may in turn produce effects on secondary target genes. Thus, identifying primary ARGs would provide insight into signal transduction pathways and transcriptional regulatory networks modulated by AR. Moreover, comparison of AR target genes in two androgen-responsive cell lines, one originating from normal prostate epithelium and the other from prostate carcinoma, might define cell-specific ARGs and suggest possible disease mechanisms.

We shall operationally define AREs as chromosomal regions that recruit AR in a native chromosomal context and modulate transcription in a reporter gene context in response to androgens. Traditionally, experimental searches for AREs have focused on short genomic segments just upstream of the transcription start sites of individual genes (Lorenzi et al. 1988; Crossley et al. 1992; Murtha et al. 1993; Rennie et al. 1993; Cleutjens et al. 1996); computational approaches have used predefined ARBS motifs to query these same regions, albeit with only modest success (Nelson et al. 2002; Steketee et al. 2004; Magee et al. 2006).

Here we have used RNA expression profiling to identify ARGs, and chromatin immunoprecipitation together with microarray analysis (ChIP–chip) to define a subset of primary ARGs, thus localizing putative AREs to be validated in reporter assays. Rather than focusing primarily on promoter proximal regions (Phuc Le et al. 2005) or CpG islands (Weinmann et al. 2002) or individual chromosomes (Carroll et al. 2005), we developed tiled-oligonucleotide microarrays spanning ∼104 kb centered on the predicted transcription start site of each of several hundred candidate hormone-responsive genes. We assessed the relationship between receptor occupancy and hormone responsive expression, characterized the genomic organization of AR-regulated genes and AREs, surveyed the complexity and composition of the AREs, and examined the physiologic roles of the primary ARGs.

Results

Identification of androgen-responsive genes

The HPr-1AR cell line, derived from immortalized normal human prostate epithelium, is androgen-responsive and expresses wild-type AR; androgen promotes differentiation of these cells, reminiscent of the normal actions of androgen on the prostate (Ling et al. 2001). We treated HPr-1AR cultures with synthetic androgen R1881, or vehicle as a control, for 4, 8, 15, and 24 h, and identified androgen-regulated transcripts on spotted cDNA microarrays. Of 157 putative ARGs identified by this analysis, 71 were repressed and 86 were induced [Fig. 1A; Supplemental Fig. 1]. To validate the expression microarray data, primers were designed for 53 putatively repressed and 72 putatively induced genes, and real-time quantitative polymerase chain reaction (qPCR) subsequently established that 42 were repressed [Fig. 1B] and 66 were induced [Fig. 1C]. These 108 differentially expressed genes included 13 (12%) previously identified ARGs, such as SGK (DePrimo et al. 2002; Nelson et al. 2002), as well as 95 (88%) ARGs not previously implicated as androgen targets.

Identification of AR binding regions

To identify genomic regions occupied by AR, HPr-1AR cells were treated with R1881 or vehicle for 4 h, a time by which AR occupancy was significantly enriched at target genes such as SGK and FKBP5 (E.C. Bolton and K.R. Yamamoto, unpubl.). Following AR ChIP, immunoprecipitated DNA fragments were amplified, labeled, and hybridized to custom NimbleGen tiled-oligonucleotide microarrays. We interrogated ∼104 kb genomic regions centered on the transcription start sites of 548 candidate hormone-responsive genes to identify AREs and to determine the distribution of AREs at promoter proximal, distal, and intragenic regions. These human candidate genes included our 157 putative and validated ARGs identified in HPr-1AR cells, 241 putative and validated glucocorticoid receptor (GR) responsive genes identified in A549 lung adenocarcinoma cells and U2OS-GR osteosarcoma cells (Rogatsky et al. 2003; Wang et al. 2004; So et al. 2007), and 150 genes potentially hormone-responsive in other cell types (e.g., 35 ARGs from LNCaP cells; DePrimo et al. 2002; Nelson et al. 2002). GR, which is closely related to AR, binds in vitro with similar affinity as AR to consensus sequences in the elementary half-sites GGTACAnnnTGTTCT (Claessens et al. 2001). Across the ∼57 Mb of DNA interrogated, 593 additional “bystander” genes were incidentally included, giving a total of 1141 genes examined in our tiled microarray (genes assigned by NCBI Build 35, hg17).

The AR ChIP–chip hybridizations directly compared DNA enriched in androgen-treated and control samples. To reduce the rate of stochastic false positives, we intersected putative AR binding peak data for duplicate independent experiments, in which 778 and 714 peaks were detected, with 509 AR binding peaks identified in both replicates. Primers for qPCR were designed for 67 AR binding peaks, and conventional AR ChIP was performed in quadruplicate experiments to establish the remaining false-positive fraction (1 of 64 or <2%), resulting in 508 androgen-stimulated AR binding regions (ARBs); <10 of these 508 ARBs may be false positives. Sixteen additional peaks identified in only one of the AR ChIP–chip replicates were also included following validation in quadruplicate experiments by conventional AR ChIP, resulting in 524 ARBs [Supplemental Table 1].

ARB validation and detection of additional ARGs

To illustrate our validation process for ChIP–chip data, we describe our analysis of an ARBR 1.2 kb upstream of the SGK gene. In this case, a functional GRE was previously known to reside in this area (Itani et al. 2002). In addition, we mapped an androgen-stimulated DNase I
Figure 1. qPCR validation of the expression of ARGs identified by microarray. (A) Most of the 157 putative ARGs identified by expression microarray analysis were subsequently validated using qPCR. The colorimetric representation shows genes, indicated by HUGO Gene Nomenclature Committee gene symbols, whose transcripts were repressed (B, green) and induced (C, red) by androgen. The color intensity reflects the relative fold change in transcript level for androgen- versus vehicle-treated cells. The time course spanned 4, 8, 15, and 24 h. The mean change in expression level for genes above the orange line was greater than twofold across the time course, whereas the mean change in expression for those genes below the line was greater than 1.5 fold.

Of the 524 ARBRs we identified, 206 ARBRs (39% of ARBRs) were within ~50 kb of the transcription start site of a validated ARG identified by our expression microarray analysis, whereas most of the remaining 318 ARBRs (61% of ARBRs) were located near genes absent from the expression microarray. We tested by qPCR 189 of these novel ARBR-proximal genes and found that 22 were androgen-repressed and 75 were androgen-induced (Fig. 3A–C). Of these 97 additional differentially expressed genes, only 10 (10%) were known ARGs (DePrimo et al. 2002; Nelson et al. 2002).

We then sought to determine which of the 108 validated ARs are also putative primary ARGs by analyzing the overlap between the ARBR and ARG data sets; we found that 75 of the 108 ARGs were near ARBRs. From the 524 ARBRs and all 205 differentially expressed ARGs, we identified 352 ARBRs (67% of ARBRs) near 172 primary ARGs (84% of ARGs) (Fig. 3D). Thus, AR occupancy of ARBRs correlates closely with androgen-responsive expression of linked genes; indeed, we sug-

hypersensitive site to this region (Supplemental Fig. 2).

Finally, our ChIP–chip and conventional ChIP experiments (Fig. 2A) revealed AR occupancy in this same region (ARBR6.35) as well as in regions 1.6 kb (intron 5) and 11.9 kb (3' of SGK gene) downstream from the SGK transcription initiation site, defining these regions as ARBRs.

As with SGK, we observed ARBRs near ARGs [within ~50 kb upstream of or downstream from the transcription start site], including FKBPS, ELL2, TMEPAI, UAP1, and FN1, all of which are androgen-responsive in LNCaP cells [DePrimo et al. 2002; Nelson et al. 2002]. ARBRs were also observed near genes not previously implicated as androgen targets, including TIPARP [Fig. 2B]. Conventional AR ChIP was performed on a representative subset of ARBRs (Fig. 2C), and >98% were confirmed in this more rigorous test. Taken together, these findings validated the approach used in this study to identify authentic ARBRs and AR regulated genes and confirmed these regions as bona fide vivo AR targets.
suggest that ChIP–chip is a powerful tool for target gene discovery that is distinct from and complementary to expression microarray analysis. For the remaining 33

![Figure 2](genesdev.cshlp.org)  
**Figure 2.** ARBRs: identification by ChIP–chip and validation by conventional ChIP. The visual representation of ARBRs on chromosomes 6 (A) and 3 (B) are shown. ARBRs are shown as equilateral red bars immediately below the chromosomes. Expanded views of (A) SGK and (B) TIPARP gene loci are shown based on the May 2004 human genome freeze hg17 in the UCSC browser using GenBank RefSeq positions. For SGK (3′→5′ orientation) and TIPARP (5′→3′ orientation), the ChIP–chip signals (red bars) are shown for duplicate experiments and indicate the relative fold enrichment of AR-immunoprecipitated DNA fragments for androgen- versus vehicle-treated cells. The ChIP scanning [Wang et al. 2004] signals (red circles) demonstrate validation of the ChIP–chip data and show the relative fold enrichment of AR ChIP for androgen- versus vehicle-treated cells. Data represent the mean ± SD, n = 4. (C) Additional ARBRs (numbered) identified by ChIP–chip were validated using conventional AR ChIP. The relative fold enrichment of AR ChIP is shown for androgen- versus vehicle-treated cells; data represent the mean ± SD, n = 4. The ARBRs found near or within genes that are repressed [green], induced [red], or unresponsive [black] following androgen treatment are indicated. A region near the HSPA1A gene, which is not occupied by AR, is shown as a control. Compared to the control, all ARBRs demonstrate significant AR occupancy, P < 0.05.

ARGs [16% of ARGs], our failure to detect nearby ARBRs may reflect (1) peaks below the threshold set in the AR ChIP–chip analysis; (2) ARBRs outside of the ∼104-kb windows surveyed here; or (3) ARGs that are secondary rather than primary targets of AR regulation. To test whether ARBRs are enriched near ARGs, and to assess AR ChIP–chip selectivity, we determined the mean frequency of ARBR occurrence (Table 1). For the 56,950 kb of DNA scanned, we identified 524 ARBRs, or one ARBR per 109 kb of genomic DNA. Of the 18,150 kb of DNA containing the 172 primary ARGs, we found 352 (67%) ARBRs, i.e., one ARBR per 52 kb of genomic DNA. For the remaining 38,800 kb of DNA scanned, even if all of the remaining 172 (33%) ARBRs were adjacent to genes whose expression is androgen-independent, a maximum of one ARBR per 226 kb was detected. We conclude that, relative to the microarray gene population, biased as it was toward hormone-responsive genes, ARBRs were enriched near ARGs (Wilcoxon rank sum test, P < 10^{-10}), consistent with the ChIP–chip data and the notion that AR occupancy strongly correlates with AR-regulated expression.

**Primary ARG clusters**

Intersection analysis of the ARG and ARBR data revealed two AR-regulated gene clusters consisting of multiple ARBRs and ARGs. An 800-kb region on chromosome 4 contained seven validated ARBRs, seven repressed ARGs and one expressed but androgen-unresponsive gene [Fig. 4A]. Similarly, a 370-kb region on chromosome 1 contained 13 validated ARBRs, 12 induced ARGs, and 12 androgen-unresponsive genes [Fig. 4B]. We are intrigued that AR regulation within these ARG clusters is negative or positive but not both, implying functional coordination resulting in part from genomic organization.

Interestingly, each ARG cluster contains genes of related function: The androgen-repressed ARGs on chromosome 4 include IL8, CXCL1, CXCL2, and CXCL3, which encode CXC chemokines involved in angiogenesis [Belperio et al. 2000]. Also on chromosome 4, EPGN, EREG, and AREG encode epithelial growth factor [EGF]-like mitogens that stimulate MAP kinase signaling [Kochupurakkal et al. 2005]. Within the induced cluster on chromosome 1, two ephrin genes, EFNA4 and EFNA3, are AR-regulated. Ephrin receptors are receptor-protein tyrosine kinases involved in cell migration and angiogenesis [Holder and Klein 1999]. Thus, the AR-regulated genes within these clusters are involved in key cellular processes, dysregulation of which may influence cancer progression.

**ARBRs display ARE activities**

To initially assess whether our ARBRs are indeed functional AREs, we cloned 22 500-bp fragments, each containing an ARBR found within or near an androgen-induced gene, into a reporter plasmid, upstream of a minimal E4 TATA sequence [Lin et al. 1988] driving the expression of luciferase. HPr-1AR cells transfected with
individual ARBR reporter constructs were treated with R1881 or vehicle control for ∼20 h, which is sufficient for robust hormone-dependent changes in luciferase activity for constructs containing AREs. Twenty [91%] of the 22 tested ARBRs displayed significant androgen-induced luciferase activity [Fig. 5A,B]. Moreover, mutation of the putative ARBSs contained within the ARBRs abolished nearly all of the androgen-induced luciferase activity [Fig. 5B]. These data demonstrate that AR occupies ARBRs by ARBS recognition, and that the vast majority of the ARBRs tested conferred AR regulation in a reporter context. We conclude that ARBRs correspond in general to functional AREs at primary ARGs, and from hereon we shall denote such ARBRs as AREs.

**ARE distribution within or near ARGs**

We sought to determine the location of AREs relative to the transcription start site of the closest ARGs; as 6% of our ARGs are annotated with two or more start sites, we independently analyzed the distribution of the AREs relative to the longest [Fig. 6A] and shortest [Supplemental Fig. 3] transcripts. For the 172 ARGs near AREs, only ∼33% of AREs (32% relative to the longest variants and 33% relative to the shortest variants) resided within 10 kb of the predicted transcription start site; most were located downstream from the ARG transcription start site. Similarly, for the 67% of AREs found >10 kb from the ARG transcription start site, most were situated downstream. Relative to the substructure of the primary ARGs, 43% of AREs were intragenic [exonic and intronic], 15% were downstream [Fig. 6B], and 42% were upstream; among ARGs with upstream AREs, only 28% were within 5 kb of the transcription start site. We examined 154 of the 185 genes containing intragenic ARBRs and found that 92 [60%] were ARGs, whereas the expression of 62 [40%] of these genes was independent of androgen in HPr-1AR. Taken together, our data indicate...
Table 1.  AR ChIP–chip selectivity for HPr-1AR binding sequences for AP-1, retinoic acid receptor (RAR), FAC (Matys et al. 2003), we concluded that potential pairing these sequence motifs with matrices in TRANS-

| Region | Size (kb) | Total ARBRs | ARBRs/100kb | ARBR |  
|--------|-----------|-------------|-------------|------|  
| Total  | 56,950    | 524         | 0.9 ± 1.2   | 109  |  
| Near 205 validated ARGs | 21,547 | 352 | 1.6 ± 1.3 | 61 |  
| Near 172 primary ARGs | 18,150 | 352 | 1.9 ± 1.2 | 52 |  
| Near genes unresponsive to androgen | 28,421 | 68 | 0.2 ± 0.6 | 418 |  
| Near remaining genes | 6,982 | 104 | 1.5 ± 0.9 | 67 |  
| Excluding regions near 205 validated ARGs | 35,403 | 172 | 0.5 ± 0.8 | 206 |  
| Excluding regions near 172 primary ARGs | 38,800 | 172 | 0.4 ± 0.8 | 226 |  

aThe mean ± SD ARBRs per 100 kb for each region.  
bThe androgen responsiveness of these genes was not determined.

that most ARGs identified by expression profiling are primary ARGs, but that most ARGs lack promoter-proximal AREs.

Interestingly, ARE density was significantly lower for androgen-repressed genes [1.5 ± 1.1 AREs/ARG, mean ± SD] than for induced genes [2.7 ± 1.8 AREs/ARG, Wilcoxon rank sum test, P < 10-6]. In addition, repressed ARGs contained intragenic AREs [19 of 65 AREs, 29%] at a lower frequency than did induced genes [156 of 340 AREs, 46%], whereas a larger fraction of AREs resided downstream from repressed ARGs [20 of 65 AREs, 31%] compared to induced genes [40 of 340 AREs, 12%] [Fig. 6B]. Conceivably, then, hormonal activation or repression of an ARG may relate in some way to ARE position and the number of AREs per ARG, implying that it may be important for mechanistic studies to identify multiple AREs near primary AR target genes. We speculate that androgen-mediated transcriptional regulation may commonly employ combinatorial coordination of multiple AREs located upstream of, within, and downstream from an ARG, e.g., SGK [Fig. 2A] and TIPARP [Fig. 2B].

Potential composite AREs

Although relatively few native HREs have been examined, many of these appear to be composite elements, with binding sequences for two or more factors defining platforms for regulatory complex assembly (Lefstin and Yamamoto 1998; Carroll et al. 2005). We employed unbiased searches using BioProspector [Liu et al. 2002] or MobyDick [Bussemaker et al. 2000; So et al. 2007] to identify recurring sequence motifs among our 524 AREs [500-bp sequences centered at the AR binding peak] [Fig. 7]. As expected, the most prominent motifs were the partially palindromic ARBS itself, occurring in 69% of AREs (Fig. 7A), and the TGTTCCT half-site of the ARBS [Fig. 7B]. In addition, we identified other recurring motifs, with Bonferroni corrected P < 10-4, by comparing these sequence motifs with matrices in TRANS-FAC [Matys et al. 2003], we concluded that potential binding sequences for AP-1, retinoic acid receptor [RAR], zinc finger protein 42 [ZNF42], hepatocyte nuclear factor 4 receptor α [HNF-4α], and early growth response [EGR] factor [Fig. 7B] were detected in >142 [27%] of our 524 AREs.

Detection of additional motifs within AREs suggested potential points of regulatory cross-talk and may facilitate analysis of combinatorial regulation at the molecular level. Indeed, several of the identified sequence motifs appear from other studies to be functional. For example, our computational analysis revealed an AP-1 binding motif at ARE4.13 of the interleukin-8 [IL8] gene, which Freund et al. [2004] showed to be important for AP-1-mediated transcriptional regulation of IL8 promoter expression. We suggest that many AREs are likely composite elements, and that regulatory cross-talk at AREs may typically modulate transcription of ARGs.

ARGs modulate diverse biological processes

To infer biological processes regulated by androgens and AR, the 205 validated ARGs were surveyed by shared Gene Ontology [GO] Biological Processes data using the Database for Annotation, Visualization and Integrated

Figure 4.  Intersection of ARBR and ARG data reveals primary ARG clusters.  [A] An 800-kb region on chromosome 4 containing seven validated ARBRs [red asterisks] was found to contain genes that were repressed [green] or unresponsive [black] following treatment with androgen.  [B] A 370-kb region on chromosome 1 containing 13 validated ARBRs [red asterisks] was found to contain genes that were induced [red] or unresponsive [black] following treatment with androgen. Gene loci are shown based on the May 2004 human genome freeze hg17 in the UCSC browser using GenBank RefSeq positions. The androgen responsiveness of some genes was not determined [gray]. Genes above the heavy black line are transcribed from left to right whereas those below the line are transcribed from right to left. The longest transcriptional variants for each gene are represented. The ChIP–chip signals [red bars] are shown for duplicate experiments and indicate the relative fold enrichment of AR-immuno-precipitated DNA fragments for androgen- versus vehicle-treated cells.

2010 GENES & DEVELOPMENT

Cold Spring Harbor Laboratory Press on July 20, 2018 - Published by genesdev.cshlp.org Downloaded from
Discovery (DAVID) 2006 (Dennis et al. 2003). The enriched GO Biological Process categories (Table 2) were dominated by ARGs whose products control cell proliferation and differentiation, cell–cell signaling, and morphogenesis, consistent with the known physiologic roles of androgens as modulators of cellular proliferation, communication, and development. We found it striking that seven ARGs whose products function in steroid metabolism were induced by androgens (Table 3), suggesting that some of the resulting lipid products may modulate cellular responses to steroids, or exert other signaling effects.

Comparing the HPr-1AR ARGs to those previously reported in LNCaP prostate carcinoma cells, we found that only 23 [11%] of the 205 HPr-1AR ARGs were also androgen-responsive in LNCaP cells (Table 4; DePrimo et al. 2002; Nelson et al. 2002). Remarkably, nine (39%) of these 23 ARGs were differentially regulated, i.e., androgen-induced in one cell line and androgen-repressed in the other. Moreover, 17 [74%] of these 23 ARGs were found to contain nearby AREs, and therefore are primary AR targets.

These results underscore the strong cell specificity of AR function and suggest that, in principle, extending these analyses of primary AR-regulated genes could reveal the androgen-responsive gene networks that modulate prostate cell proliferation, communication, differentiation, and possibly cancer progression.

**Discussion**

We found that multiple genes involved in prostate cell growth, proliferation, and perhaps cancer progression are regulated directly by AR. AREs were detected near 84% of ARGs, suggesting that, during our relatively short times of hormone exposure, most ARGs in HPr-1AR cells are primary targets of AR rather than secondary consequences of hormone exposure, and that AR occupancy is a major determinant of AR responsiveness. It will be interesting to extend our cell culture analysis to assess the relative roles for primary and secondary ARGs in developing and adult prostate tissue, and during prostate cancer progression.

**Roles of ARGs in biological processes**

Consistent with the physiologic roles of androgens as modulators of cellular proliferation, signaling, and develop-
chain acyl CoA synthetase ($\text{ACSL1}$), which activates long-chain fatty acids for $\beta$-oxidation or acylation by cholesterol acyl transferase ($\text{SOAT2}$). $\text{SOAT2}$ in turn converts cholesterol, from which steroids are derived, into cholesterol esters for storage. Together, these primary ARGs expand the AR signaling network to include specific lipid metabolism pathways; some of the resulting lipid products may modulate cellular responses to steroids or exert other signaling effects.

### Table 2. Summary of Gene Ontology (GO) categories

| GO biological processes                              | $P$-values$^a$ |
|------------------------------------------------------|---------------|
| Cell proliferation                                   | 0.00019       |
| Regulation of cell proliferation                     | 0.00007       |
| Positive regulation of cell proliferation            | 0.00041       |
| Negative regulation of cell proliferation            | 0.0023        |
| Cell communication                                   | 0.00065       |
| Cell–cell signaling                                  | 0.0085        |
| Signal transduction                                  | 0.0065        |
| Cell surface receptor linked signal transduction     | 0.0025        |
| Regulation of signal transduction                    | 0.052         |
| Development                                           | 0.0011        |
| Tissue development                                   | 0.024         |
| Mesoderm development                                 | 0.044         |
| Cell differentiation                                 | 0.0067        |
| Morphogenesis                                         | 0.0028        |
| Organ morphogenesis                                   | 0.065         |
| Cellular morphogenesis                                | 0.18          |
| Regulation of cell size                              | 0.067         |
| Regulation of cell growth                            | 0.031         |
| Response to stimulus                                 | 0.0011        |
| Defense response                                      | 0.011         |
| Response to chemical stimulus                         | 0.013         |
| Response to stress                                    | 0.0068        |
| Response to external stimulus                         | 0.00036       |
| Taxis                                                | 0.0072        |
| Response to wounding                                  | 0.00073       |
| Wound healing                                         | 0.0044        |
| Cellular lipid metabolism                             | 0.22          |
| Lipid biosynthesis                                    | 0.21          |
| Steroid metabolism                                   | 0.057         |
| Steroid biosynthesis                                  | 0.081         |
| Sterol metabolism                                    | 0.085         |
| Sterol biosynthesis                                   | 0.078         |
| Cholesterol metabolism                               | 0.066         |
| Cholesterol biosynthesis                              | 0.055         |
| Negative regulation of metabolism                    | 0.011         |
| Negative regulation of cellular metabolism            | 0.012         |
| Sodium ion transport                                  | 0.040         |
| Negative regulation of progression through cell cycle | 0.045         |
| Negative regulation of transcription                 | 0.046         |

$^a$ $P$-values were determined using as background the 11,222 predicted protein-coding genes that we interrogated.

Figure 7. ARBSs and other cis-regulatory motifs are enriched within AREs. Unbiased motif searches of all 524 AREs revealed the presence of significantly recurring motifs. [A] BioProspector analysis identified a consensus ARBS, which is visually represented in Weblogo [http://weblogo.berkeley.edu]. [B] Moby-Dick analysis detected a consensus AR half-site motif as well as putative motifs for AP-1, RAR, ZNF42, HNF-4α, and EGR. These enriched sequence motifs are represented using IUPAC symbols [http://www.iupac.org]. Motifs where the putative regulatory factors were not determined (ND) are indicated. P-value calculations are described in Materials and Methods.
In general, we were struck by the remarkable degree of cell and gene specificity inherent in androgen regulation, as 182 (89%) of the 205 ARGs we identified were apparently androgen-unresponsive in LNCaP cells (DePrimo et al. 2002; Nelson et al. 2002). Similarly, comparisons of glucocorticoid-responsive genes in different cell types (Rogatsky et al. 2003; Wang et al. 2004; Phuc Le et al. 2005; So et al. 2007) revealed little overlap. It will be important in future work to assemble expression and response data from multiple androgen-responsive cell types to better define and understand tissue-specific transcriptional networks radiating outward from AR.

### Potential composite AREs

In 69% of AREs (Fig. 7A), we readily identified a partially palindromic ARBS motif, and mutation of the motif virtually abolished the ARE activities (Fig. 5B). The remaining AREs lacked a statistically significant consensus ARBS motif and might be “tethering elements”, at which AR associates through protein–protein interactions with DNA-bound nonreceptor regulatory factors, rather than itself binding to DNA (Murtha et al. 1997; Sato et al. 1997; Lefstin and Yamamoto 1998). In any case, our approach was broadly effective in identifying functional AREs that contain or lack consensus ARBSs, and that either promote or repress AR-regulated transcription.

Combinations of cis-regulatory motifs can function as composite response elements to integrate multiple signals and to elicit distinct regulatory effects in different cellular or physiologic contexts. These biological nodes present opportunities to define regulatory interactions at the molecular and functional levels. As a first step, we scanned for additional sequence motifs within the AREs and identified several significantly recurring motifs, including putative sites for AP-1, RAR, ZNF42, HNF-4α, and EGR. We note that AP-1 motifs are known to operate at several ARGs (Murtha et al. 1997; Sato et al. 1997; Chen et al. 2006), and that increases in AR and the AP-1 factor c-Jun correlate with prostate cancer progression (LaTulippe et al. 2002; Tiniakos et al. 2006). Assessing regulatory cross-talk between AR and AP-1 signaling at AREs will thus be informative but technically challenging as the dimeric AP-1 factors are drawn from a superfamily of genes [c-Jun, JunB, and JunD] that can homo- or heterodimerize with c-Fos, FosB, Fra-1, and Fra-2.

### ARE distribution, combinatorial regulation, and ARG clusters

Studies of promoter proximal AREs have yielded an emerging view of coordinated recruitment of AR, nuclear coregulators, chromatin remodeling factors, and the transcriptional machinery (Shang et al. 2002; Metzger et al. 2005; Wang et al. 2005; Yamane et al. 2006). Whether such findings present a complete picture, however, is a question brought into focus by results showing that most AREs are quite remote from ARGs. We suggest that some remote AREs may act analogously to the β-globin locus control region (LCR), which exerts temporal control over expression of genes within the β-globin gene cluster from 60 kb away (Lawrence 2003; Takayama et al. 2007) hypothesize that distal intergenic as well as intragenic AREs function to regulate primary ARGs.

We identified 172 primary AR target genes and 524 AREs. In agreement with a recent ChIP–chip study of 10 androgen-responsive genes in LNCaP (Takayama et al. 2007), we found that most AREs were >10 kb upstream of and, or more typically, downstream from the ARG transcription start sites. As functional enhancers, AREs can operate relatively remote from promoters, and thus provide little selective pressure to assume more promoter proximal positions. Our results parallel those from ChIP–chip studies of other transcriptional regulators [Martone et al. 2003; Euskirchen et al. 2004; So et al. 2007]. We also commonly observed multiple AREs per ARG (Figs. 2, 4, 5), suggesting combinatorial regulation of individual primary ARGs.

Interestingly, 31 (15%) ARGs were themselves linked in gene clusters, consisting of ARBRs and multiple ARGs. In addition to the two large ARG clusters described here (Fig. 4A,B), we also identified five small ARG clusters consisting of at least one intergenic ARE flanked by two ARGs (E.C. Bolton and K.R. Yamamoto, unpubl.). Strikingly, all of the ARGs within any given cluster were similarly regulated, i.e., either induced or repressed by androgens, but not a mixture of both. However, some genes within a given cluster nonetheless escape AR-mediated regulation in HPt-1AR cells. Conceivably, AR-mediated regulation within ARG clusters may progress through discrete stages during male development, analogous to the temporal regulation of the β-globin locus (Lawrence 2003).
The challenges associated with testing the coordination of several AREs spanning ∼100 kb for even a single ARG are daunting. One possible approach for future work is to use integrated bacterial artificial chromosome-based reporter genes in mouse genetic studies (DiLeone et al. 2000; Dodou et al. 2004). In any case, it seems likely that intricate collaborations of multiple AREs may orchestrate recruitment of coregulatory complexes that dictate the spatiotemporal expression of AR target genes.

The present study advances our understanding of the organization and function of AREs and nearby ARGs, and also illuminates new unknowns, such as the mechanisms that restrict AR occupancy to a subset of potential AREs in a given cell context, or the contributions of individual AREs to ARG regulation, especially in the context of ARG clusters. Addressing these and other challenges raised by our study will continue to provide new insights into the mechanisms of transcriptional regulation of genes and gene networks by AR and other transcriptional regulators.

**Materials and methods**

*Cell culture, plasmids, transfections, and reporter assays*

HPr-1AR cells, which express wild-type AR, were grown as described (Ling et al. 2001) in keratinocyte serum free medium (Cat. # 17005–042, Invitrogen) and a 5% CO2 atmosphere. Plasmid pGL4.10-E4TATA (provided by Yuriy Shostak, University of California at San Francisco, San Francisco, CA) and pBEC22 (Renilla luciferase vector) were created by insertion of a 50-bp minimal E4 TATA promoter sequence (Lin et al. 1988), driving luciferase expression, into the BglII to HindIII sites of vectors pGL4.10 and pGL4.70 (Promega), respectively. Individual ∼500-bp ARBR fragments were amplified from human genomic DNA by PCR (primer sequences in Supplemental Table 2) and inserted into the KpnI to XhoI sites upstream of the E4 TATA promoter. The QuikChange kit (Stratagene) was used to mutate two to four of the underlined positions (GGAACAnnnTGTNCN) of putative ARBSs contained within the ARBRs (primer sequences in Supplemental Table 2). Mutation of either of two consensus ARBSs within ARBR7.08 reduced androgen-induced luciferase activity [E.C. Bolton and K.R. Yamamoto, unpubl.], and mutation of both ARBSs (mARBR7.08) abolished it. For ARBR6.35 and mARBR6.35, DNA linkers containing...
only the 15-bp ARBS and mutated ARBS, respectively, were inserted into the Kpn I to Xho I sites of pGL4.10-E4TATA. HPr-1AR cells were grown in 48-well plates, cotransfected with 150 ng of reporter constructs and 150 ng of pBEC22 [Renilla luciferase vector] using Gene Juice [Qiagen]. After overnight transfection, cells were treated with 0.1% ethanol vehicle or 1 nM R1881 [New England Nuclear] for ~20 h, and luciferase activities were measured using the Dual-Luciferase Reporter Assay (Promega) and an Ultra Evolution plate reader (Tecan).

RNA isolation, reverse transcription, and qPCR
Total RNA was isolated from HPr-1AR cells using QiAshredder and RNeasy Mini kit [Qiagen]. Random-primed cDNA was prepared from total RNA using the ProtoScript First Strand cDNA kit [New England Biolabs]. cDNA was used in 50-µL qPCR reactions containing 1 unit of AmpliTaq Gold DNA polymerase (Applied Biosystems), 2.5 mM MgCl2, 500 nM of each primer (primer sequences in Supplemental Table 3), 150 µM dNTP mix, 0.2× SYBR green I dye [Invitrogen], and 1× ROX reference dye [Invitrogen] in 1× PCR Buffer II (Applied Biosystems). Primers were designed using Primer3 [http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi], and those that efficiently amplified Nuclei were collected by centrifugation (500 g for 5 min at 4°C) and resuspended in 1 mL of ice-cold IP lysis buffer (50 mM HEPES-KOH, pH 7.4, 1 mM EDTA, 150 mM NaCl, 10% glycerol, 0.5% Triton X-100, supplemented with protease inhibitors), and harvested by scraping; cell lysis continued as cells were mixed at 4°C.

For the ChIP-chip microarray, ~104 kb human genomic regions centered on the transcription start sites of 548 candidate hormone-responsive genes were tiled with isothermal 50-mer oligos that were spaced ~54 bp apart on repeat-masked DNA. Where ~104-kb regions overlapped, the surrounding genomic regions were tiled further, bidirectionally. Human DNA sequences from NCBI Build 35 [hg17] were retrieved using the UCSC Genome Browser [http://genome.ucsc.edu]. NimbleGen Systems [http://www.nimblegen.com] labeled the LMPCR-amplified AR ChIP samples from vehicle and R1881-treated HPr-1AR cells with Cy3 and Cy5, respectively, hybridized labeled samples to the custom microarrays, and measured relative signal intensities. Androgen-enriched peaks were detected by Mpeak [http://www.stat.ucla.edu/~zml/mpeak; Kim et al. 2005] using default parameters with the following modifications: threshold prefilter = 2(mean) + 4(SD); local maximum = 350 bp.

Computational analysis
For analysis of enriched motifs, human repeat-masked DNA sequences from NCBI Build 35 [hg17] were retrieved using the UCSC Genome Browser [http://genome.ucsc.edu]. BioProspector analysis [Liu et al. 2002] was initially performed using nucleotide widths (w) 12, 14, and 16, and 18 on ARBSs to identify ARBSs, and the top motifs were masked to identify other motifs. For MobyDick analysis [Bussemaker et al. 2000], human and human/mouse aligned sequences (UCSC Genome Browser) were used as inputs to identify enriched motifs [So et al. 2007]. Similar motifs were clustered and Bonferroni corrected P-values of enrichment were determined as described [So et al. 2007] (see Supplemental Materials and Methods). The top BioProspector w14 position weight matrix (PWM) was used to score ARBSs with a false-positive rate of <10%. This upper bound was calculated from randomly sampling unbound regions (Supplemental Fig. 4).

For other motifs with P < 10−2, PWMs were constructed and compared to known binding sites in TRANSFAC professional v.9.3 [Matys et al. 2003]. To associate motifs with putative regulatory factors using relative entropy (Kullback-Liebler divergence) with a cutoff <6.0.

Acknowledgments
We thank the members of the Yamamoto lab for discussions and reagents, and Hee Jung Chung, Mark Diamond, Miles Pfall, Nirao Shah, Sebastiaan Meijing, and Tony Gerber for insightful comments on the manuscript. Cell line HPr-1AR and plasmid pGL4.10-E4TATA were generous gifts from Patrick Ming-Tat Ling [University of Hong Kong] and Yuriy Shostak, respectively. This work was supported by NRSA postdoctoral fellowship F32DK65402 to E.C.B., predoctoral fellowship 10GB-0104 of the University of California Breast Cancer Research Program to A.Y.S., NSF predoctoral fellowship to C.C., Packard Fellowship in Science and Engineering to H.L.; and research grants from NIH [CA101042 to C.M.H., GM070808 to H.L., CA020535 to K.R.Y.].

References
Belpietro, J.A., Kcane, M.P., Arenberg, D.A., Addison, C.L., Ehler, J.E., Burdick, M.D., and Strieter, R.M. 2000. CXC chemokines in angiogenesis. J. Leukoc. Biol. 68: 1–8.
Regulation of AR target genes

by protein-protein interaction between the androgen receptor and AP-1/c-Jun in the human prostate cancer cell line LNCaP. J. Biol. Chem. 272: 17485–17494.

Shang, Y., Myers, N., and Brown, M. 2002. Formation of the androgen receptor transcription complex. Mol. Cell 9: 601–610.

Shibata, N., Arita, M., Misaki, Y., Dohmec, N., Takio, K., Ono, T., Inoue, K., and Arai, H. 2001. Supersaturant protein factor, which stimulates the conversion of squalene to lanosterol, is a cytosolic squalene transfer protein and enhances cholesterol biosynthesis. Proc. Natl. Acad. Sci. 98: 2244–2249.

So, A.Y., Chaiuvorapul, C., Bolton, E.C., Li, H., and Yamamoto, K.R. 2007. Determinants of cell- and gene-specific transcriptional regulation by the glucocorticoid receptor. PLoS Genet. 3: e94. doi: 10.1371/journal.pgen.0030094.

Steketee, K., Ziel-van der Made, A.C., van der Korput, H.A., Houtsmuller, A.B., and Trapman, J. 2004. A bioinformatics-based functional analysis shows that the specifically androgen-regulated gene SARG contains an active direct repeat androgen response element in the first intron. J. Mol. Endocrinol. 33: 477–491.

Takayama, K., Kaneshiko, K., Tsutsu, S., Horie-Inoue, K., Ikeda, K., Urano, T., Ijichi, N., Ouchi, Y., Shiraighe, K., Aburatani, H., et al. 2007. Identification of novel androgen response genes in prostate cancer cells by coupling chromatin immunoprecipitation and genomic microarray analysis. Oncogene 26: 4453–4463.

Tiniakos, D.G., Mitropoulos, D., Kyroudi-Voulgari, A., Soura, K., and Kittas, C. 2006. Expression of c-jun oncogene in hyperplastic and carcinomatous human prostate. Urology 67: 204–208.

Wang, J.C., Derynck, M.K., Nonaka, D.F., Khodabakhsh, D.B., Haq, C., and Yamamoto, K.R. 2004. Chromatin immunoprecipitation (ChIP) scanning identifies primary glucocorticoid receptor target genes. Proc. Natl. Acad. Sci. 101: 15603–15608.

Wang, Q., Carroll, J.S., and Brown, M. 2005. Spatial and temporal recruitment of androgen receptor and its coactivators involves chromosomal looping and polymerase tracking. Mol. Cell 19: 631–642.

Waterham, H.R. 2006. Defects of cholesterol biosynthesis. FEBS Lett. 580: 5442–5449.

Weinmann, A.S., Yan, P.S., Oberley, M.J., Huang, T.H., and Farnham, P.J. 2002. Isolating human transcription factor targets by coupling chromatin immunoprecipitation and Cpg island microarray analysis. Genes & Dev. 16: 235–244.

Yaman, K., Tomazou, C., Tsukada, Y., Erdjuman-Bromage, H., Tempst, P., Wong, J., and Zhang, Y. 2006. JHDM2A, a JmjC-containing H3K9 demethylase, facilitates transcription activation by androgen receptor. Cell 125: 483–495.

Biotechnol. 20: 835–839.

Lorenzi, F., Stavenhagen, I., Kalff, M., and Robins, D.M. 1988. A complex androgen-responsive enhancer resides 2 kilobases upstream of the human Slp gene. Mol. Cell. Biol. 8: 2350–2360.

Lubahn, D.B., Joseph, D.R., Sullivan, P.M., Willard, H.F., French, F.S., and Wilson, E.M. 1988. Cloning of human androgen receptor complementary DNA and localization to the X chromosome. Science 240: 327–330.

Magee, J.A., Chang, L.W., Stormo, G.D., and Milbrandt, J. 2006. Direct, androgen receptor-mediated regulation of the FKBP5 gene via a distal enhancer element. Endocrinology 147: 590–598.

Martone, R., Euskirchen, G., Bertone, P., Hartman, S., Royce, T.E., Luscombe, N.M., Rinn, J.L., Nelson, F.K., Miller, P., Gerstein, M., et al. 2003. Distribution of NF-κB-binding sites across human chromosome 22. Proc. Natl. Acad. Sci. 100: 12247–12252.

Matys, V., Fricke, E., Geffers, R., Gossling, E., Haubrock, M., Hehl, R., Hornischer, K., Karas, D., Kel, A.E., Kel-Margoulis, O.V., et al. 2003. TRANSFAC: Transcriptional regulation, from patterns to profiles. Nucleic Acids Res. 31: 374–378. doi: 10.1093/nar/gkf108.

Metzger, E., Wissmann, M., Yin, N., Muller, J.M., Schneider, R., Peters, A.H., Gunther, T., Buettner, R., and Schule, R. 2005. LSD1 demethylates repressive histone marks to promote androgen-receptor-dependent transcription. Nature 437: 436–439.

Murtha, P., Tindall, D.J., and Young, C.Y. 1993. Androgen induction of a human prostate-specific kallihekine, hKLK2: Characterization of an androgen response element in the 5′ promoter region of the gene. Biochemistry 32: 6459–6464.

Murtha, P.E., Zhu, W., Zhang, J., Zhang, S., and Young, C.Y. 1997. Effects of Ca2+ mobilization on expression of androgen-regulated genes: Interference with androgen receptor-mediated transactivation by AP-1 proteins. Prostate 33: 264–270.

Nelson, P.S., Clegg, N., Arnold, H., Ferguson, C., Bonham, M., White, J., Hood, L., and Lin, B. 2002. The program of androgen-responsive genes in neoplastic prostate epithelium. Proc. Natl. Acad. Sci. 99: 11890–11895.

Oliver, L.M., Chambless, K.L., Gibson, K.M., and Krisans, S.K. 1999. Characterization of phosphohemovaleon kinase: Chromosomal localization, regulation, and subcellular targeting. J. Lipid Res. 40: 672–679.

Panet-Raymond, V., Gottlieb, B., Bertel, L.K., Pinsky, L., and Trifiro, M.A. 2000. Interactions between androgen and estrogen receptors and the effects on their transactivational properties. Mol. Cell. Endocrinol. 167: 139–150.

Phuc Le, P., Friedman, J.R., Schug, J., Brestelli, J.E., Parker, J.B., Bochkis, I.M., and Kaestner, K.H. 2005. Glucocorticoid receptor-dependent gene regulatory networks. PLoS Genet. 1: e16. doi: 10.1371/journal.pgen.0010016.

Rennie, P.S., Bruchovsky, N., Leco, K.J., Sheppard, P.C., McQueen, S.A., Cheng, H., Snoek, R., Hamel, A., Bock, M.E., MacDonald, B.S., et al. 1993. Characterization of two cis-acting DNA elements involved in the androgen regulation of the probasin gene. Mol. Endocrinol. 7: 23–36.

Rogatsky, I., Wang, J.C., Derynck, M.K., Nonaka, D.F., Khodabakhsh, D.B., Haq, C.M., Darimont, B.D., Garabedian, M.J., and Yamamoto, K.R. 2003. Target-specific utilization of transcriptional regulatory surfaces by the glucocorticoid receptor. Proc. Natl. Acad. Sci. 100: 13845–13850.

Sato, N., Sada, M.D., Bruchovsky, N., Saatcioglu, F., Rennie, P.S., Sato, S., Lange, P.H., and Cleave, M.E. 1997. Androgenic induction of prostate-specific antigen gene is repressed...
Cell- and gene-specific regulation of primary target genes by the androgen receptor

Eric C. Bolton, Alex Y. So, Christina Chaivorapol, et al.

*Genes Dev.* 2007, 21:
Access the most recent version at doi:10.1101/gad.1564207

Supplemental Material

http://genesdev.cshlp.org/content/suppl/2007/08/06/21.16.2005.DC1

References

This article cites 59 articles, 26 of which can be accessed free at:
http://genesdev.cshlp.org/content/21/16/2005.full.html#ref-list-1

License

Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or click here.