Chapter 5
The Functionality of Prostate Cancer Predisposition Risk Regions Is Revealed by AR Enhancers

Houtan Noushmehr, Simon G. Coetzee, Suhn K. Rhie, Chunli Yan, and Gerhard A. Coetzee

Abstract Prostate Cancer (PCa) genetic risk has recently been defined in numerous genome-wide association studies (GWAS), which revealed more than 50 disease-associated single nucleotide polymorphisms (SNPs), known as tagSNPs, each at a different locus. More than 80% of these tagSNPs are located in noncoding regions of the genome for which functionality remains unknown. We and others hypothesize that at least some of these SNPs affect noncoding genomic regulatory signatures such as enhancers. Many research laboratories including ours have profiled the genomic

H. Noushmehr
Norris Cancer Center, Keck School of Medicine, University of Southern California, Los Angeles, CA 90033, USA

Department of Preventive Medicine, Keck School of Medicine, University of Souther California, Los Angeles, CA 90033, USA

Department of Urology, Keck School of Medicine, University of Southern California, Los Angeles, CA 90033, USA

Department of Genetics, Medical School of Ribeirão Preto, University of São Paulo, Ribeirão Preto, São Paulo 14049-900, Brazil

Epigenome Center, Keck School of Medicine, University of Southern California, Los Angeles, CA 90033, USA

S.G. Coetzee
Norris Cancer Center, Keck School of Medicine, University of Southern California, Los Angeles, CA 90033, USA

Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA 90033, USA

Department of Urology, Keck School of Medicine, University of Southern California, Los Angeles, CA 90033, USA

Department of Genetics, Medical School of Ribeirão Preto, University of São Paulo, Ribeirão Preto, São Paulo 14049-900, Brazil

distribution of androgen receptor (AR) and the dynamic state of the PCa genome for active chromatin regions (H3K9,14ac), open chromatin regions (DNaseI), enhancers (H3K4me1/2), and active/engaged enhancers (H3K27ac). In order to identify candidate functional SNPs, which may confer risk associated with PCa, we recently developed an open-source (R/Bioconductor) package, called FunciSNP (Functional Integration of SNPs), which systematically integrates SNPs from the 1000 genomes project with these biologically active chromatin features. Here we report several potential AR enhancers, defined by genome-wide data and from chromatin biofeatures, which may be functionally involved in PCa risk.

Keywords Enhancer • Androgen Receptor Occupied Regions (ARORs) • Chromatin • Genome • Single Nucleotide Polymorphism • Post-GWAS function

Abbreviations

AR	Androgen receptor
DHT	Dihydrotestosterone
ARE	Anrogen responsive element
ARORs	Androgen receptor occupied regions
GWAS	Genome-wide association studies
1000gp	1000 genomes project
SNP	Single nucleotide polymorphisms
FunciSNP	Functional Identification of SNPs
LD	Linkage disequilibrium

S.K. Rhie • C. Yan
Norris Cancer Center, Keck School of Medicine, University of Southern California,
Los Angeles, CA 90033, USA

Department of Preventive Medicine, Keck School of Medicine, University
of Southern California, Los Angeles, CA 90033, USA

Department of Urology, Keck School of Medicine, University of Southern California,
Los Angeles, CA 90033, USA

G.A. Coetzee (✉)
Norris Cancer Center, Keck School of Medicine, University of Southern California,
Los Angeles, CA 90033, USA

Department of Preventive Medicine, Keck School of Medicine, University
of Southern California, Los Angeles, CA 90033, USA

Department of Urology, Keck School of Medicine, University of Southern California,
Los Angeles, CA 90033, USA

USC/Norris Cancer Center, NOR 6411, MS#73, 1441 Eastlake Ave.,
Los Angeles, CA 90089, USA
e-mail: coetzee@usc.edu

5.1 Introduction

Within the past 5 years, prostate cancer (PCa) genome-wide association studies (GWAS) have identified more than 50 risk loci, with more than 80% found in non-coding genomic regions (Fig. 5.1) [1–4]. Integrating data associated with genomic profiles of androgen receptor (AR) occupancy and other enhancers to these known risk regions will inevitably provide an informative genomic landscape that may yield insight and understanding into the etiology and ultimately treatment of PCa. Characterizing and understanding the role of the AR in prostate cancer and the dynamic interactions between AR and nucleosomes is an important area of PCa research [5, 6] and consequently is hypothesized to play "driver" roles at GWAS risk loci. The aim of this chapter is to describe recent research in profiling the AR interactions with genomic features. In addition, we will present and describe the use of a novel bioinformatic tool (developed by our group), which integrates several large-scale data types to identify well-demarcated new genomic risk regions at known risk loci that may have important implications for PCa genetic predisposition.

5.1.1 Normal Function of AR in Prostate

Understanding the role of AR in PCa and the regulation of AR activity remains an intense and important area of research. In normal prostate development and maintenance, AR drives differentiation and regulates the expression of select genes for the production of proteins that mainly function in seminal fluid development, similar to the role estrogen receptor provides in the regulation and production of milk in the breast [7]. Therefore, AR is not normally classified as an oncogene during prostate development and maintenance, whereas its role in prostate tumorigenesis is indisputably as a driver and as such may be viewed as an oncogene during the progression of the disease.

5.1.2 AR Significance in PCa

As early as the 1940s, Dr. Charles B. Huggins (a Nobel laureate) discovered that steroid hormones such as androgen could be used to control the spread of prostate and breast cancer. Orchiectomy, the removal of the male testes, was found to be an effective surgical procedure for the treatment of advanced PCa, later replaced with GnRH agonist [8, 9]. In the 1950s, it was found that patients who relapse after orchiectomy respond better to andrenalectomy, which indicates that the tumors may be stimulated by residual androgens from adrenal glands [8, 9]. From the mid-1990s, multiple two-arm trials looked at the effects of castration versus castration+AR antagonist (such as flutamide, milutamide, cyproterone acetate, and
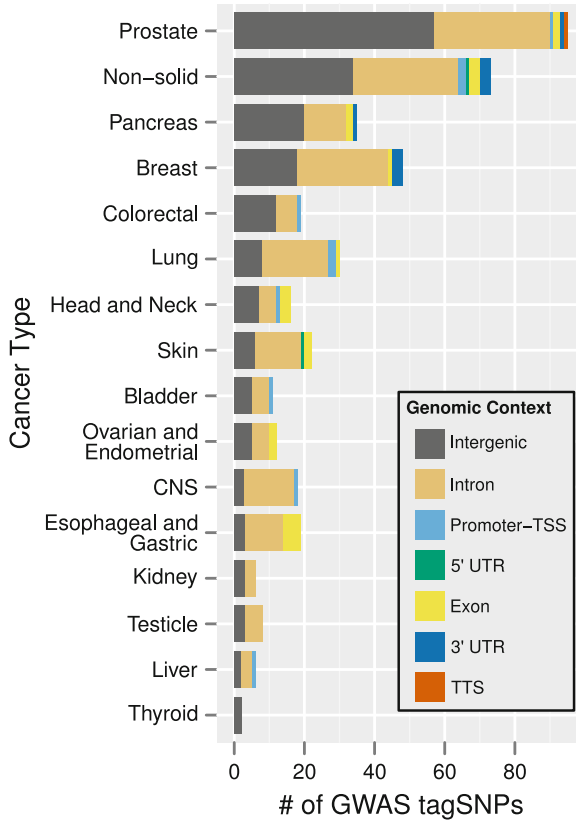
Fig. 5.1 Summary of all known GWAS SNPs across a number of different cancer types. GWAS SNPs were annotated using known genomic features supplied by HOMER (version 3.9) [43], using build hg19 as reference. GWAS SNPs were extracted from the UCSC Genome table browser, track name "gwasCatalog" with a *P*-value cut-off of 9e−06 [2]. Figure and legend reproduced from Coetzee SG. et al. 2012 [3] with permission from Oxford University Press

bicalutamide) to block effects of residual androgen expression [8, 9]. The hypothesis at the time was to further block AR, which would provide an alternative to castration procedures. After hundreds of millions of dollars invested into researching the effect of these antagonists, it was found that there was minimal improvement, with modest responses to AR antagonists given at relapse after castration, with a general conclusion that more complete AR blockade was no more effective then castration alone [9].

During the same time, basic research labs found that AR and AR regulated genes were highly expressed in castration recurrent/resistant prostate cancer (CRPC). The AR gene was found to be amplified in ~30% of CRPC and AR antagonists selected for AR mutations [7, 10]. By 2000, work by several research groups found that high

intra-tumoral androgen levels in CRPC were synthesized from weak adrenal androgens, and this led to the hypothesis that androgens were being developed in a de novo fashion from cholesterol products [9, 11, 12]. These observations provided important clinical application, and as recent as 2011, survival advantage from post- and pre-chemotherapy for Abiraterone (CYP17A1 inhibitor, an enzyme for androgen synthesis) in CRPC was found to decrease serum testosterone from ~20–50 ng/dl to <1 ng/dl [11, 12]. These clinical studies provided evidence that AR to some degree is driving PCa progression [11, 12]. In 2012, survival advantage for a novel AR antagonist, MDV3100, was found to be highly selective [13–15]. Despite progressive AR decreasing during tumor progression, tumors have found many genetic means of maintaining AR mediated signaling. We know that most tumors that have become castration resistant remain dependent on AR mediated activity, which has been shown with the advancement of drugs such as MDV3100, which continue to target AR pathway axis [10, 15]. Even more recently, exome sequencing of 50 CRPCs revealed many somatic mutations in the androgen-signaling pathway, DNA repair, and histone/chromatin modifier genes [16]. These and other important findings shed both clinical and biological importance in determining the mechanisms mediating intrinsic and acquired resistance to androgen and AR mediated pathways.

5.1.3 Advancement in Sequencing Technology Allows for Comprehensive Profiling of Specific Genomic and Epigenomic Marks

The last 15 years of biomedical research has been fueled in part by development of new technologies, techniques, and protocols. The recent advent of next-generation sequencing has not only spawned many new exciting applications and opportunities but also challenges for cancer researchers. ChIP-seq is a technique to systematically detect genome-wide protein:DNA interactions and histone modifications. In a given landscape of the genome, there are a variety of diverse proteins that interact with our genome in a sequence and tissue-specific manner. During a typical ChIP-seq experiment, the cells are treated with formaldehyde to stably cross-link all protein:DNA complexes. The chromatin is then sheared and fragmented to approximately 200–1,000 bp by sonication. Generally, such an experiment is done with about 2 million cells to enrich the amount of protein:DNA complex using a specific antibody targeting a protein of interest. Once the purified and enriched DNA fragments are obtained, high-throughput sequencing is performed and fragments of 50–100mers are mapped back to the reference genome. Currently, deep sequencing technology can sequence up to 100 million reads for a given experiment, and this number is very likely to increase exponentially while the overall cost of sequencing will drop over the next 5–10 years [17, 18]. Analyzing deep sequencing data for a particular ChIP experiment requires several important steps, which include initial quality control, peak

calling, peak quality assessment, motif analysis, and large prediction and annotations of the data [17]. There are many different peak calling algorithms available, including MACS2 and FindPeaks [17].

In Sect. 5.2, we will describe the current understanding of AR occupancy regions (ARORs) in prostate cancer cell lines and associated enhancer elements in promoters and non-promoters of genes profiled using deep sequencing technologies. Section 5.3 will introduce the current state of GWAS in PCa and the idea of a novel bioinformatic tool we developed to integrate data profiled from deep sequencing, GWAS, and the 1000 genomes project to identify candidate regulatory elements involved in PCa progression and/or risk.

5.2 AROR Identification and Genome-wide Distribution of Enhancer Elements

Myles Brown and his research group was among the first to systematically identify thousands of AR binding sites in PCa cells after DHT treatment by AR chromatin immunoprecipitation (ChIP) followed chip hybridization (initially) or by deep sequencing (more recently) (ChIP-seq; described in detail in Sect. 5.1.3) [19]. In addition, we and others profiled AROR in DHT-treated LNCaP cells. We reported that about 20% of histone H3 acetylated at Lys-9 and Lys-14 (H3K9,14ac) enrichment surrounds ARORs [20]. We also observed a high number of AROR peaks distributed in noncoding regions, while a significant enrichment of ARORs was found in promoters of known genes only with overlapping H3K9,14ac (Fig. 5.2a, b). If ARORs overlapping and not overlapping known H3K9,14ac peaks are separated, we observe that the genomic distribution to all known transcription start sites are significantly depleted at promoters for AROR peaks not overlapping H3K9,14ac (Fig. 5.2c, d). Interestingly, AROR peaks overlapping H3K9,14ac are enriched at promoters (Fig. 5.2e). This genomic profile of AROR has given researchers an unprecedented view of the wide distribution of AR binding to the human genome in PCa after DHT stimulation. It has been determined that AR binding sites have AR element (ARE) motif and are generally associated with a consensus motif for FOXA1 [19, 21]. In our own study, we performed de novo motif analysis on our identified AROR peaks. Significantly identified AROR peaks are highly enriched for the known ARE motif, while ARORs located in distant or non-promoters were also enriched for the FOXA1 motif (Fig. 5.3). It has been shown that FOXA1 binds with AR and other steroid receptors in a diverse set of tissues [19]. Among other factors, FOXA1 has been shown to function as a pioneer factor to open the locus prior to AR binding [22]. Further work by the Brown lab and others have found that sites occupied by FOXA1 (and the recruitment of FOXA1) are highly associated with enhancers, since these regions usually have nucleosomes with enhancer histone marks such as H3K4me1 and H3K4me2 [19, 20, 23] (Table 5.1). Prior to androgen stimulation and AR binding, ARE is covered with loosely associated nucleosome that contains a variant of
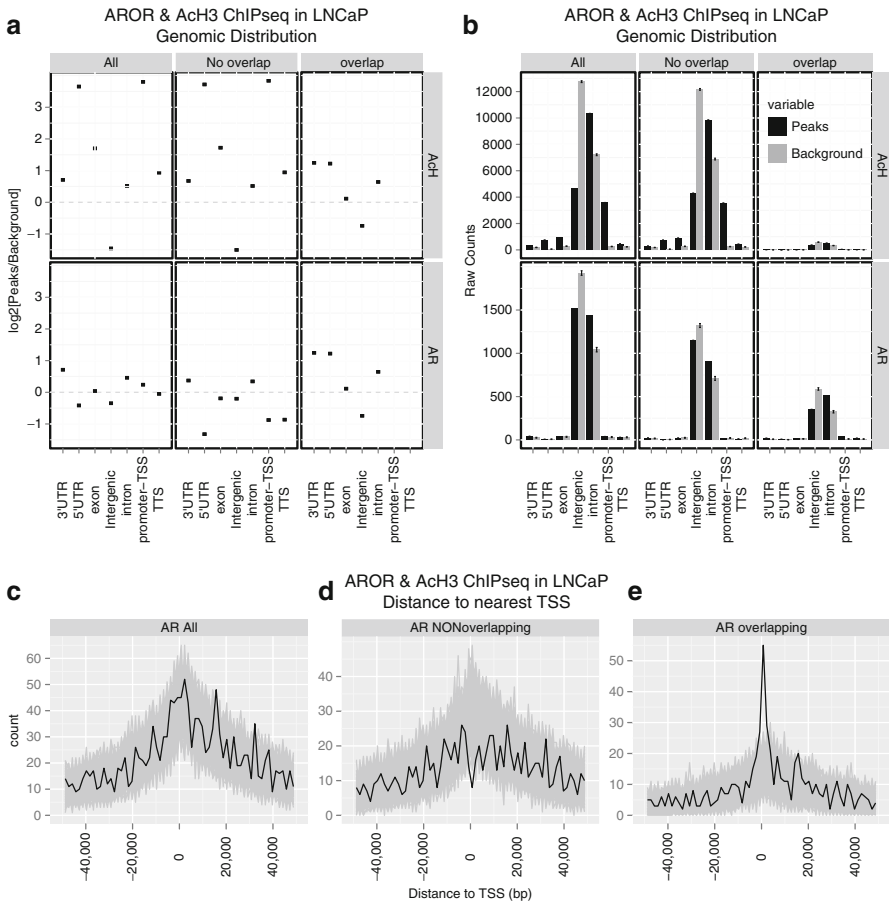
Fig. 5.2 Complete genomic distribution of AROR and H3K9,14ac. (**a–b**) raw counts and log2 ratio (peaks versus background) for H3K9,14ac (AcH) and AR distribution across known genomic regions, 3′UTR, 5′UTR, exon, intergenic, intron, promoter-TSS (promoter transcription start site), TTS (transcription termination site). Each column of plots indicates differing sets of peaks, All— total peaks per type (AcH or AR), No overlap (distinct peak files), overlap (peak files in common between AcH and AR). *Black* color indicates observed peaks, while *gray* colors indicate background as computed 1,000 times (randomly generated regions selected genome wide). (**c–e**) Genomic profile of AROR as a function of distance to the nearest canonical gene's TSS. ARORs were divided into (**c**) total, (**d**) AROR not overlapping AcH, (**e**) AROR overlapping AcH. *Red* color indicates observed peaks; while *gray* colors indicate background as computed 1,000 times (randomly generated regions selected genome wide)

H2A.Z which is displaced by AR [24]. At the chromatin level, AR was found enriched at ARE sites associated with H3K4me1 and H3K4me2. These distinct epigenetic marks are highly associated with FOXA1, which provides recruitment in CRPC [25]. This process is reversed by LSD1 (lysine-specific demethylase 1), which mediates demethylation of H3K4me1 and H3K4me2 [25]. Thus it appears the sequence of

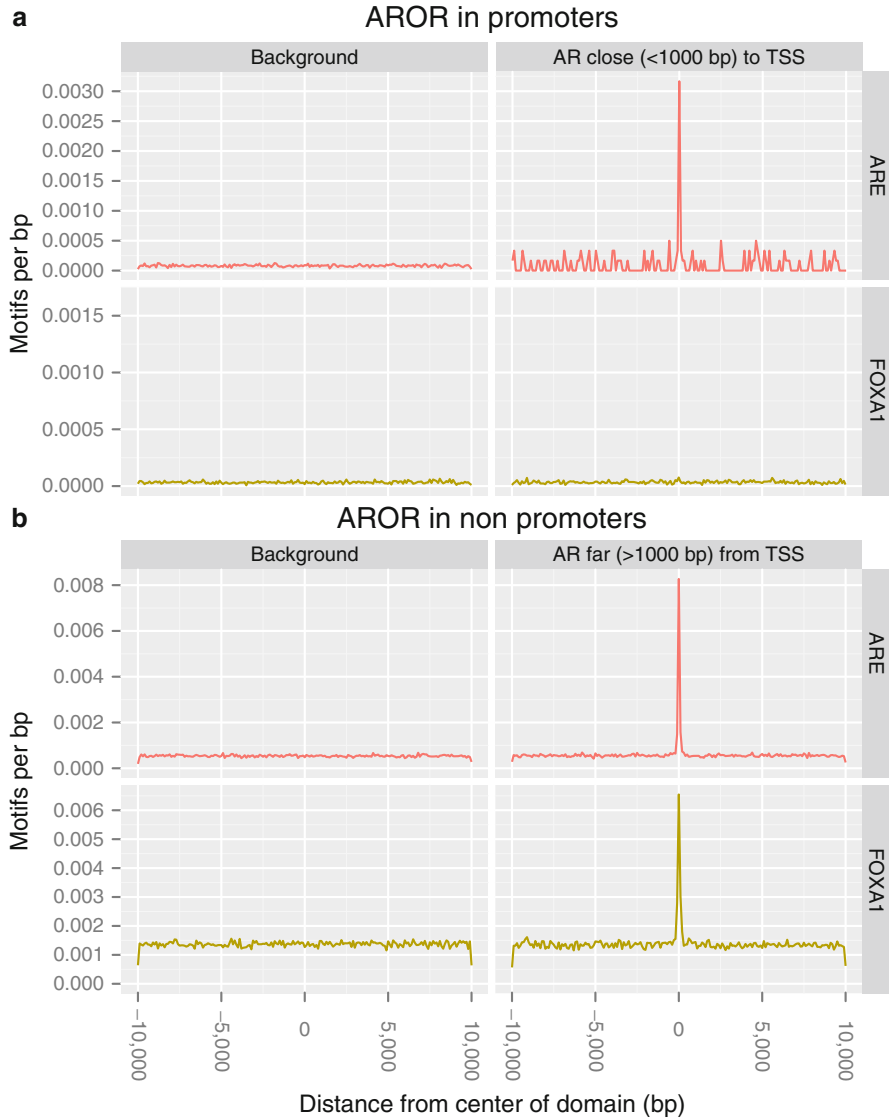**Fig. 5.3** De novo motif analysis results of ARORs in promoters (<1 kbp from TSS) and non-promoters (>1 kbp from TSS). (**a–b**) Genomic profiles are generated using ARORs centered on top two de novo motifs as identified by the ARORs in non-promoters. (**c**) Enrichment table and motif sequence are described
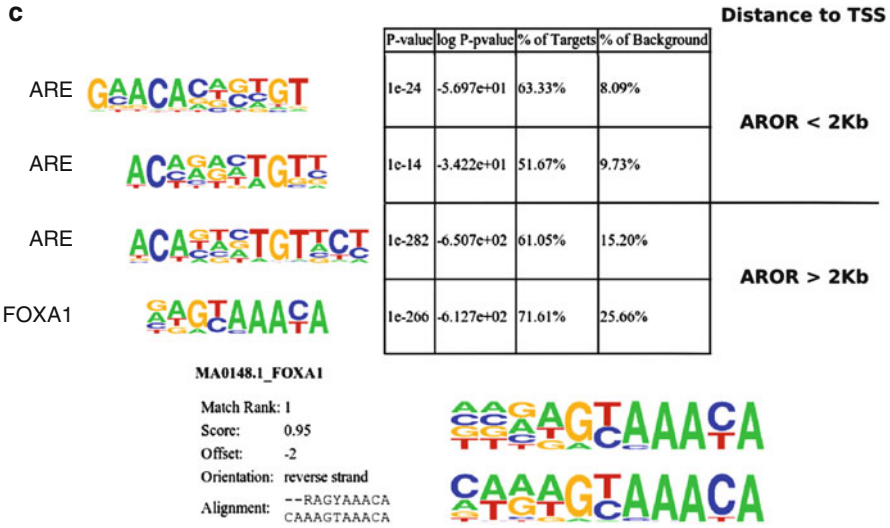
c

		P-value	log P-pvalue	% of Targets	% of Background	
ARE		1e-24	-5.697e+01	63.33%	8.09%	**Distance to TSS**
ARE		1e-14	-3.422e+01	51.67%	9.73%	**AROR < 2Kb**
ARE		1e-282	-6.507e+02	61.05%	15.20%	**AROR > 2Kb**
FOXA1		1e-266	-6.127e+02	71.61%	25.66%	

MA0148.1_FOXA1

Match Rank: 1
Score: 0.95
Offset: -2
Orientation: reverse strand
Alignment: --RAGYAAACA
                CAAAGTAAACA

Fig. 5.3 (continued)

Table 5.1 Diverse biological features compiled from several different sources. All biofeatures were profiled in LNCaP cells treated with DHT

Biofeatures	Biological features	Num peaks
AR [20]	AROR	3,100
AR [26]	AROR	6,214
AR [22]	AROR	205
AR+DNaseI	AROR + Open chromatin	1,055
AR+H3K9,14ac [20]	AROR + Active/engaged enhancers	580
AR+HK4me1	AROR + Enhancers	456
DNAseI+H3K4me1+H3K27ac	Open + Active/Enhancers	5,174
H3K27ac [22]	Active/engaged	12,578
H3K4me1 [22]	Enhancers	19,149
H3K9/14ac [20]	Open chromatin	956
Genomic features (hg19 human build)		
Exons		698,453
CTCF Only [30]		9,916
DNAseI Only [30]		175,507
DNAseI + CTCF [30]		4,977
Known promoters (−1000 to +100)		39,700

events start with the epigenetic alterations of histones, specifically methylation of K4 on the H3 complex, and recruitment of FOXA1, thereby allowing for an open conformational change of the nucleosome, thereby allowing for poised interactions for AR stimulated by DHT to incorporate with ARE, and thereby activate targeted genes responsible for progression and differentiation of the prostate.

Further subsequent findings in advanced castration resistant PCa using LNCaP-abl cells have shown that AR regulates the expression of specific G2/M genes (CDK1, UBE2C, others) in advanced castration resistant but not in androgen sensitive LNCaP cells [23]. In summary, the spectrum of AR regulated genes may change during tumor development and progression.

Recently, it was discovered that FOXA1 suppressive sites are enriched for insulator protein binding sites, which are sites of compact chromatin (CTCF-FOXA1-Groucho family complexes) [26, 27]. They observed, by knocking down FOXA1 followed by ChIPseq of AR binding sites, hundreds of AREs are dependent of FOXA1 while if FOXA1 is deleted, 3,500 AR binding sites are revealed and are independent of FOXA1. These sites are enriched for insulator protein binding sites, CTCF–FOXA1, and set of corepressive elements [26, 27]. The overall physiological importance is unclear, but these studies illustrate that the dynamic integrity of AR binding to the DNA is quite flexible and provides a complex environment with a multitude of diverse functions yet to be fully characterized.

DNaseI hypersensitivity (HS) is a method to assay chromatin accessibility [28]. It allows for a comprehensive profile of distinct genomic regions, which are preferentially unbound by nucleosomes and thus are characterized as open chromatin region. DNaseI sequencing results are aligned to the human genome reference, compared to a background control experiment and peaks are called as described in Sect. 5.1.3, to identify candidate open chromatin regions in LNCaP cells. Most studies have shown that DNaseI hypersensitive sites correlate very well with regulatory elements and enhancer activity in a variety of tissues and cell types [22, 29]. We put together a table, which summarizes the total number of AR and available enhancer peaks across a number of different studies, as well as the overlap with data downloaded from the ENCODE project [30] (Table 5.1).

5.3 Identification of Significant Risk Variants Associated with PCa

5.3.1 Statistical Analysis and Bioinformatics Tool to Identify Candidate Functional Enhancer Elements

GWAS have yielded numerous single nucleotide polymorphisms (SNPs) associated with many phenotypes. The main goals for genetic risk factor identification have been to allow clinicians and epidemiologist a factor to predict risk for a disease or treatment such as chemotherapy or surgical prevention. It has also been important to allow researchers an opportunity to understand the biology of cancer by identifying

genes or loci involved in the development of the cancer and finally to identify novel genetic regions or loci for drug targeting and treatment.

The first GWAS study published in 2005 identified a risk region for age-related macular degeneration disease [31], and GWAS have been constantly evolving and expanding their reach in large part due to the advancement of genotyping technology. During the last 6 years, it is apparent that GWAS efforts fostered by these advancements have made significant discoveries, for example, more than 1,449 tagSNPs have been published for association with well over 200 different diseases or traits at a *p*-value of 5×10^{-8} [2, 3]. It is also apparent that as new methods and stratifications by populations emerge, these studies could yield even higher numbers of variations associated to a particular disease or trait. Interestingly, approximately 200 genomic variants have been reported for more than 20 different cancer types, and it is expected that this number of associated risk loci will likely exceed 300 by the end of this year [32]. Remarkably, we and others have noted that the vast majority of these variants associated with cancer, in particular with prostate and breast, are enriched in noncoding regions (Table 5.2, Fig. 5.1) [3]. Despite the many novel and insightful biological discoveries GWAS have provided to the scientific community, we have a long way to go before we can fully understand genetic heritability and risk. In the case of breast cancer, there are very rare allele frequencies in a given population with very high effect size, many identified through family-based studies, candidate gene re-sequencing, and regions with candidate risk association. The rare variance contributes on total, over 25% to understanding the familial risk of most cancer whereas only 10–15% contributes to common risk with low to modest effect size [33]. However, the field is currently looking to see if the risk variant for low-to-common allele frequency can be identified with intermediate to large effect size, which is believed to be largely due to genetic factors. Currently, with the advancement in sequencing technology and methodology, we are at the forefront of making discoveries of these low to common variants with measurable effect size with suitable power. And as the advancement of sequencing technology increases and the cost to perform such high-throughput deep sequencing reaches a critical cost–benefit level, the major bottleneck will be computational methods and tools needed in order to tease out the genetic variants and risk associated with many of these critically debilitating human diseases.

At the core of all association studies dominated during the past 6–7 years is the idea of linkage disequilibrium (LD). LD is the observation that two or more alleles in a population segregating together more often than one would expect by chance [3, 34]. What this definition means to researchers interested in identifying genetic basis or identify candidate functional variants in a particular disease is that the identity of the functional variant is not required. As an example, a causal or putative functional variant is located on the ancestral allele, and this variant will segregate successfully through the population over time. If we are able to visualize this variant, then we can measure LD and therefore indirectly identify the variant using markers or tags surrounding the causal or putative functional variant [35, 36].

SNPs are the most common form of variation we have measured in the human population, and therefore SNP markers provide the best opportunity to identify the

Table 5.2 Reported GWAS tagSNP associated with PCa

SNP rsID	Chrom	StartPos	Genomic region	Genomic context	Nearest gene	Risk allele freq.	p-value	OR	95% CI	Reference (PMID)
rs13385191	chr02	20888265	2p24.1	Intron	C2orf43	0.44	8.00E-08	1.15	[1.10–1.21]	[44]
rs1465618	chr02	43553949	2p21	Intron	THADA	0.23	2.00E-08	1.08	[1.03–1.12]	[45]
rs6545977	chr02	63301164	2p15	Intergenic	OTX1—RPL27P5	–	5.00E-07	–	–	[45]
rs10187424	chr02	85794297	2p11.2	Intergenic	GGCX—VAMP8	0.59	3.00E-15	1.09	[1.06–1.12]	[46]
rs12621278	chr02	1733311553	2q31.1	Intron	ITGA6	0.94	9.00E-23	1.33	[1.25–1.43]	[45]
rs2292884	chr02	238443226	2q37.3	Intron	MLPH	0.25	4.00E-08	1.14	[1.09–1.19]	[47]
rs9311171	chr03	37996477	3p22.2	Intron	CTDSPL	–	2.00E-06	–	–	[48]
rs17181170	chr03	87173324	3p12.1	Intergenic	VGLL3—CHMP2B	–	3.00E-08	–	–	[45]
rs10934853	chr03	128038373	3q21.3	Intron	EEFSEC	0.28	3.00E-10	1.12	[1.08–1.16]	[49]
rs6763931	chr03	141102833	3q23	Intron	ZBTB38	0.45	2.00E-08	1.04	[1.01–1.07]	[46]
rs345013	chr03	145173788	3q24	Intergenic	C3orf58—GM2AP1	–	5.00E-06	–	–	[48]
rs10936632	chr03	170130102	3q26.2	Intergenic	SKIL—CLDN11	0.52	7.00E-22	1.11	[1.08–1.14]	[46]
rs17021918	chr04	95562877	4q22.3	Intron	PDLIM5	0.65	4.00E-15	1.11	[1.08–1.15]	[45]
rs7679673	chr04	106061534	4q24	Intergenic	RPL6P14—TET2	0.55	3.00E-14	1.1	[1.06–1.14]	[45]
rs2242652	chr05	1280028	5p15.33	Intron	TERT	0.81	3.00E-24	1.15	[1.11–1.19]	[46]
rs12653946	chr05	1895829	5p15.33	Intergenic	IRX4—IRX2	0.44	4.00E-18	1.26	[1.20–1.33]	[44]
rs2121875	chr05	44365545	5p12	Intron	FGF10	0.34	4.00E-08	1.05	[1.02–1.08]	[46]
rs4466137	chr05	82985739	5q14.3	Intron	HAPLN1	–	3.00E-06	–	–	[48]
rs37181	chr05	115630003	5q22.3	Intergenic	COMMD10	–	0.007	1.6	[1.04–1.30]	[46]
rs130067	chr06	31118511	6p21.33	Missense	CCHCR1	0.21	3.00E-08	1.05	[1.02–1.09]	[46]
rs1983891	chr06	41536427	6p21.1	Intron	FOXP4	0.41	8.00E-08	1.15	[1.09–1.21]	[44]

SNP	Chr	Position	Band	Location	Gene	Freq	p-value	OR	95% CI	Ref
rs10498792	chr06	51666631	6p12.3	Intron	PKHD1	–	3.00E-06	–	–	[48]
rs339331	chr06	117210052	6q22.1	Intron	RFX6	0.37	2.0E-12	1.22	[1.15–1.28]	[44]
rs9364554	chr06	160833664	6q25.3	Intron	SLC22A3	0.29	6.00E-10	1.17	[1.08–1.26]	[50]
rs12155172	chr07	20994491	7p15.3	Intergenic	RPS26P30—ASS1P11	0.2	9.00E-06	1.05	[1.00–1.10]	[45]
rs10486567	chr07	27976563	7p15.2	Intron	JAZF1	0.77	2.00E-06	1.12	[1.02–1.25]	[51]
rs6465657	chr07	97816327	7q21.3	Intron	LMTK2	0.46	1.05E-08	1.12	[1.05–1.20]	[45, 50]
rs1512268	chr08	23526463	8p21.2	Intergenic	SLC25A37—NKX3-1	0.45	2.00E-11	1.18	[1.14–1.22]	[44, 45]
rs10086908	chr08	128011936	8q24.21	Intergenic	POU5F1B	–		1.7	[1.39–2.07]	[52]
rs1447295	chr08	128485038	8q24.21	Intergenic	POU5F1B—MYC	0.1	6.70E-17	1.53	[1.29–1.77]	[49, 53, 54]
rs1571801	chr09	124427372	9q33.2	Intron	DAB2IP	–	0.001	1.36	[1.13–1.65]	[55]
rs10993994	chr10	51549496	10q11.23	nearGene	MSMB	0.4	1.26E-06	1.19	[1.04–1.34]	[44, 47, 50, 51]
rs4962416	chr10	126696872	10q26.13	Intron	CTBP2	0.27	2.00E-07	1.17	[1.05–1.30]	[51]
rs7127900	chr11	2233574	11p15.5	Intergenic	TH—ASCL2	0.2	3.00E-33	1.22	[1.17–1.27]	[45]
rs10896469	chr11	56309467	11q12.1	TTS	OR5M11	–		–	–	[56]
rs12418451	chr11	68935418	11q13	Intergenic	TPCN2	–	1.20E-06	–	–	[57, 58]
rs731236	chr12	48238756	12q13.11	Exon	HDAC7	–	0.05	–	–	[59]
rs10875943	chr12	49676010	12q13.12	Intergenic	TUBA1C—PRPH	0.31	7.00E-12	1.07	[1.04–1.10]	[46]
rs902774	chr12	53273904	12q13.13	Intergenic	KRT78—RPL7P41	0.17	5.00E-09	1.17	[1.11–1.24]	[47]
rs12827748	chr12	80088577	12q21.2	Intergenic	PAWR	–	0.02	1.6	[1.1–2.4]	[59]
rs9600079	chr13	73728139	13q22.1	Intergenic	FABP5L1—KLF12	0.38	3.00E-09	1.18	[1.12–1.24]	[44]

(continued)

Table 5.2 (continued)

SNP rsID	Chrom	StartPos	Genomic region	Genomic context	Nearest gene	Risk allele freq.	p-value	OR	95% CI	Reference (PMID)
rs1529276	chr13	103928007	13q33.1	Intergenic	SLC10A2—RPL7P45	–	2.00E-06	–	–	[48]
rs11649743	chr17	36074978	17q12	Intron	HNF1B	–	1.70E-09	–	–	[60]
rs7210100	chr17	47436749	17q21.33	Intron	ZNF652	0.05	3.00E-13	1.51	[1.35–1.69]	[61]
rs1859962	chr17	69108753	17q24.3	Intergenic	CALM2P1—SOX9	0.46	2.50E-07	1.24	[1.14–1.37]	[45, 47, 50]
rs8102476	chr19	38735613	19q13.2	Intergenic	DPF1—PPP1R14A	0.54	2.00E-11	1.12	[1.08–1.15]	[49]
rs887391	chr19	41985623	19q13	Intergenic	C19orf69	–	9.40E-04	–	–	[62]
rs2735839	chr19	51364623	19q13.33	Intergenic	KLK3—KLK2	0.85	2.00E-18	1.2	[1.10–1.33]	[50]
rs9623117	chr22	40452119	22q13.1	Intron	TNRC6B	0.21	5.00E-07	1.18	[1.11–1.26]	[63]
rs5759167	chr22	43500212	22q13.2	Intergenic	RPS25P10—BIK	0.53	6.00E-29	1.16	[1.14–1.20]	[45]
rs5919432	chrX	67021550	Xq12	Intergenic	AR—OPHN1	0.81	1.00E-08	1.06	[1.02–1.12]	[46]

ancestral causing or functional allele in a given population exhibiting a shared or common disease or trait. However, as the ancestral causing allele passes through time in a population (over numerous generations), the LD breaks down significantly. This breakdown in LD structure then makes identification of variant causing allele more difficult. Therefore, a systematic identification of all known SNPs or variants is required and to this end, the HapMap project evolved. The HapMap project aim was to identify SNPs using four different human populations to put together a phased haplotype block map and identifiable LD blocks of the entire human genome [37]. This project was successful at the time, identifying 4–5 million SNPs in the profiled population. Although these blocks identified regions of interest, it did not successfully identify the candidate or functional variant [38]. Therefore, the 1000 genomes project (1000 gp) emerged to tackle the issue by harnessing the obvious observation that the world's population is quite diverse than the original four populations profiled from the HapMap project [39]. The 1000 gp set out to sequence more than 1,000 individuals across a larger number of different populations. Currently, the 1000 gp have identified more than 60 million SNPs, indels, somatic mutations, and other genomic variants across the population [39]. This has now allowed us to identify a larger number of variants among a given haploblock containing the identified risk allele. Some have argued this increase in variants coupled with unidentified functional variant provides an even bigger challenge in identifying the causal or functional variant. However, we and others have noted that this increase in number of identified variants provides a deeper resolution of the heritability and thus allows us to harness more informative variants than previously identified [3, 40]. We developed a bioinformatic tool called FunciSNP (Functional Identification of SNP), which allows us to harness the information provided by the 1000 gp, the location of the previously reported risk allele (tagSNP), and the chromatin features to identify the causal and functional regions associated with PCa [3] (Fig. 5.4b) (for results, see Sect. 5.3.3–5.3.4).

5.3.2 GWAS PCa

PCa genetic risk has recently been defined in a number of GWAS, which have revealed 51 disease-associated SNPs, known as tagSNPs, each at a different locus (Table 5.2, Figs. 5.1 and 5.4a). More than 80% of these tagSNPs are in intergenic or intron regions of the genome for which functionality remains unknown. Interestingly, the tagSNPs are distributed among many chromosomes (Fig. 5.4a) with the exception being that chromosomes 1, 14, 15, 16, 18, 20, and 21 do not contain any tagSNPs. They occur at relatively gene rich areas (inner circle of Fig. 5.4a), with the notable exceptions of the tagSNPs at chromosome 3 (middle SNP), 8q, and the distal SNP at 12q. Many tagSNPs track closely with chromatin biofeatures (outer four circles of Fig. 5.4a, and see below).

Taking the 1000 genomes data into consideration, more than 10,000 correlated SNPs are revealed, which each define risk due to linkage disequilibrium (correlated
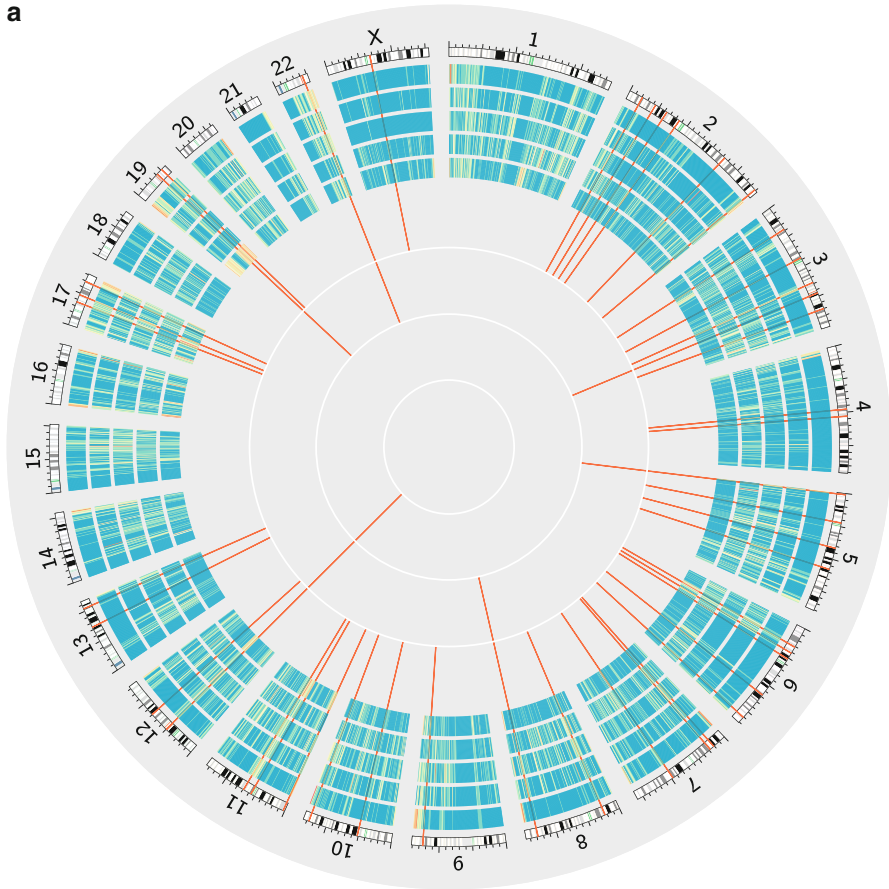
a

Fig. 5.4 GWAS integration with biological features. (**a**) Circos plot illustrating genomic distribution of the 51 known GWAS tagSNPs as of June 2012 (*red lines* indicate the position of the tagSNPs with the length reflecting the number of independent tagSNPs within a 10 MB window. The *white concentric circles* are markers for 1, 2, and 3 SNPs; outside to inside). From outermost to innermost, the colored bands describe the concentration (*red, yellow, green, blue*; high to low) DNase1 sensitivity, H3K27ac, H3K4m31, AROR, and genes, respectively. The chromatin marks were obtained from LNCaP cells. On the ideograms the locations of the centromeres are depicted in *green* and tagSNPs in *red*. (**b**) Schematic flowchart to describe FunciSNP. *Purple boxes* represent process before integration with biofeature. *Red boxes* represent information after integration with biofeature (Figure and legend reproduced from Coetzee SG. et al. 2012 [3] with permission from Oxford University Press.). (**c**) Schematic diagram indicating total number of candidate SNPs pooled from the 1000 gp that overlap one or more biological genomic features and which are in linkage disequilibrium to the original GWAS SNP. Total number of SNPs in each section is identified by a *blue box*. (**d**) Overall distribution of R^2 values for all identified 1000 gp SNPs overlapping at least one biological feature. Each identified bin reports total number of candidate 1000 gp SNPs
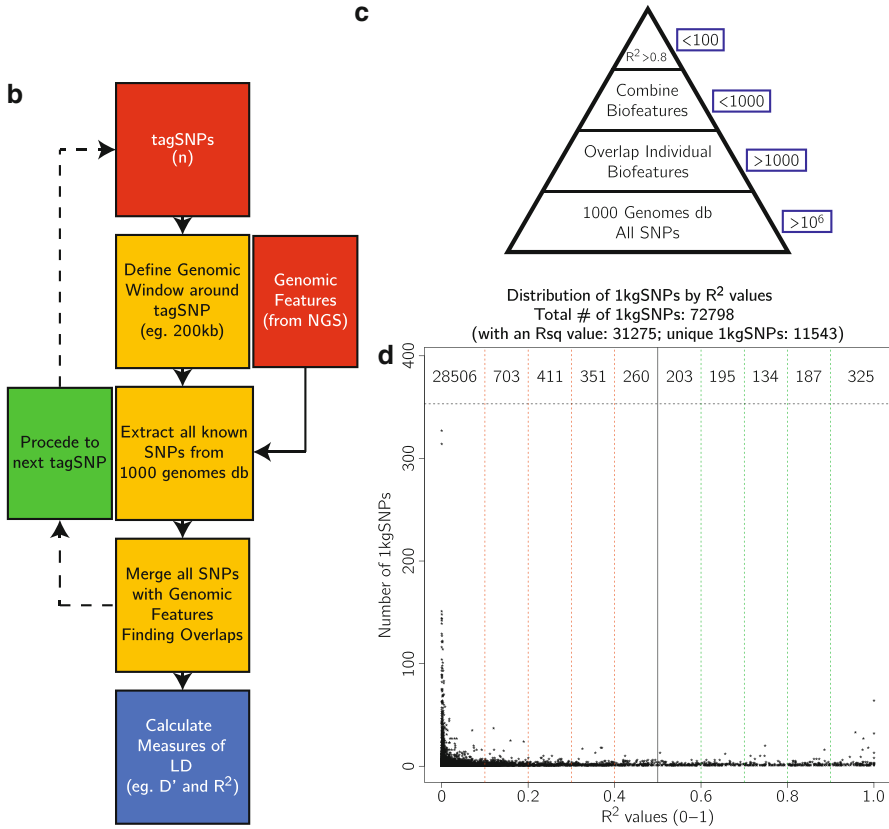
Fig. 5.4 (continued)

to the tagSNP with an $R^2 > 0.8$, Fig. 5.4c). The above therefore makes the identification of functional and/or causal SNPs not a trivial task. Therefore, we hypothesize that at least some of these SNPs affect noncoding genomic regulatory signatures, such as enhancers or insulators. As we described in Sect. 6.2, there are distinct regions in the genome profiled in PCa cells, which we and others have identified and have significant biological relevance in PCa.

5.3.3 *FunciSNP Integrates AROR, Enhancer, GWAS and 1000gp to Identify Candidate Functional Elements*

In order to reduce the number of candidate functional SNPs profiled in the 1000gp linked to the associated tagSNP in PCa, we developed an open-source (R/Bioconductor) package, called FunciSNP, which systematically integrates the

1000gp SNP data with chromatin features of interest [3]. To define functionality in noncoding DNA associated with PCa risk, we extracted open chromatin and enhancer features generated by next-generation sequencing technologies. The open chromatin state (DnaseI, H3K9,14ac), enhancers (H3K4me1), and active/engaged enhancers (H3K27ac) were generated either by our lab [20], or harvested from the ENCODE project [41], or retrieved from recent publications [22, 26] (Table 5.1 and visualized in Fig. 5.4a). All chromatin features were identified in the same PCa cell line (LNCaP). We identified 113 PCa risk correlated SNPs at androgen receptor occupied regions (ARORs), 1,545 at DNase1 sensitivity sites, and 403 at histone modified regions [H3K4me1 (160 SNPs), H3K9,14 ac (121 SNPs), H3K27ac (122 SNPs)]; all features excluded transcription start sites of known annotated genes (Figs. 5.4d, 5.5, 5.6a). ARORs coinciding with DNase1 sites revealed four novel SNPs correlated with four GWAS tagSNPs. Of the four novel surrogate SNPs, two are located 4 kb upstream from KLK3, one is located within the 3'UTR of NKX3.1, and one is located within an intron of RUVBL1 (Fig. 5.6a; http://goo.gl/cYCl7). DNase1 sites coinciding with any histone modification provided 12 novel risk regions correlated with 10 GWAS tagSNPs (Tables 5.3 and 5.4). Four are located in 8q24 genomic region and each is more than 100 kb away from a known annotated gene.

5.3.4 *Enhancer Validation in Two PCa Cell Lines and in Primary Prostate Epithelial Cells*

To measure potential enhancer activities in the eight chosen potential enhancer regions, we employed a firefly luciferase reporter (driven by a basal tk-promoter) into which ~1.2 kb potential enhancers coinciding with PCa risk correlated SNPs were cloned. These vectors, along with renilla luciferase controls, were transfected into two PCa cell lines: LNCaP cells, PC3 cells along with an AR expression vector, and normal primary prostate epithelial cells (PREC) along with an AR expression vector and stimulated with DHT. Enhancer activities were measured as previously described [42] and compared with two negative controls (regions with no enhancer histone marks) and a positive control (PSA enhancer). DHT-stimulated enhancer activity for seven of the eight potential enhancers was evident in at least one cell type (the exception is F26), indicating that the identification of enhancers using our protocol is both high and specific (Fig. 5.6b). Further analyses are required to determine the role(s) of these enhancers in PCa risk.

5.4 Future Perspective/Discussion

During the past decade three major findings in human genetics/genomics have caused a dramatic change in our appreciation of our genome and how it functions. First, the human genome project revealed that humans have only some 22,000 genes
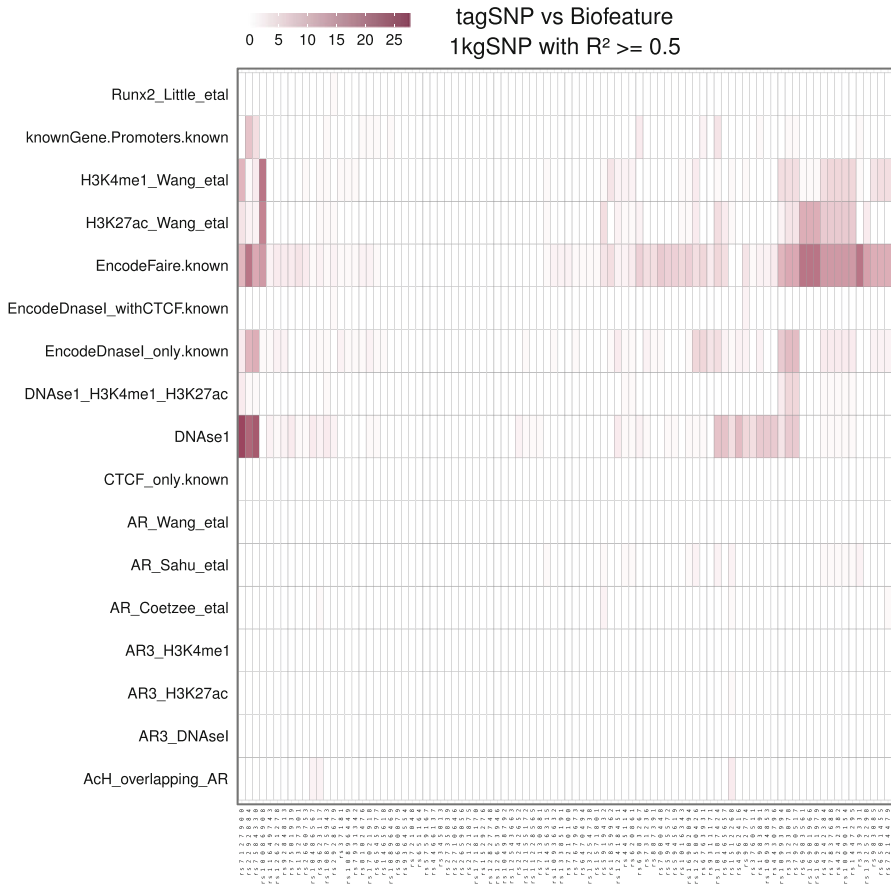
Fig. 5.5 FunciSNP heatmap of the number of 1000 gp between tagSNP and biofeature for PCa. Total number of candidate SNP is listed by color within each quadrant to represent the number of potential candidate functional SNPs overlapping a biofeature (*y*-axis), which are in linkage disequilibrium to the original GWAS risk SNP (*x*-axis)

within the 3 billion-nucleotide genome. Second, many transcription factors (such as the AR discussed above) do not preferentially occupy gene promoters preferentially but are rather scattered among the many noncoding stretches of DNA in introns and intergenic regions. Third, as pointed out above, GWAS signals for complex diseases are preferentially found in introns and intergenic regions. These three novel insights have pointed to the importance of so-called noncoding DNA (previously even referred to as junk DNA). Transcription factor occupancy, histone modification marks, and nucleosome-depleted regions in chromatin are presently used to annotate functionality within most of our genome. Ultimately this will yield important insight to understand gene expression, genetic risk of complex diseases, and genetic associations with most human phenotypes.
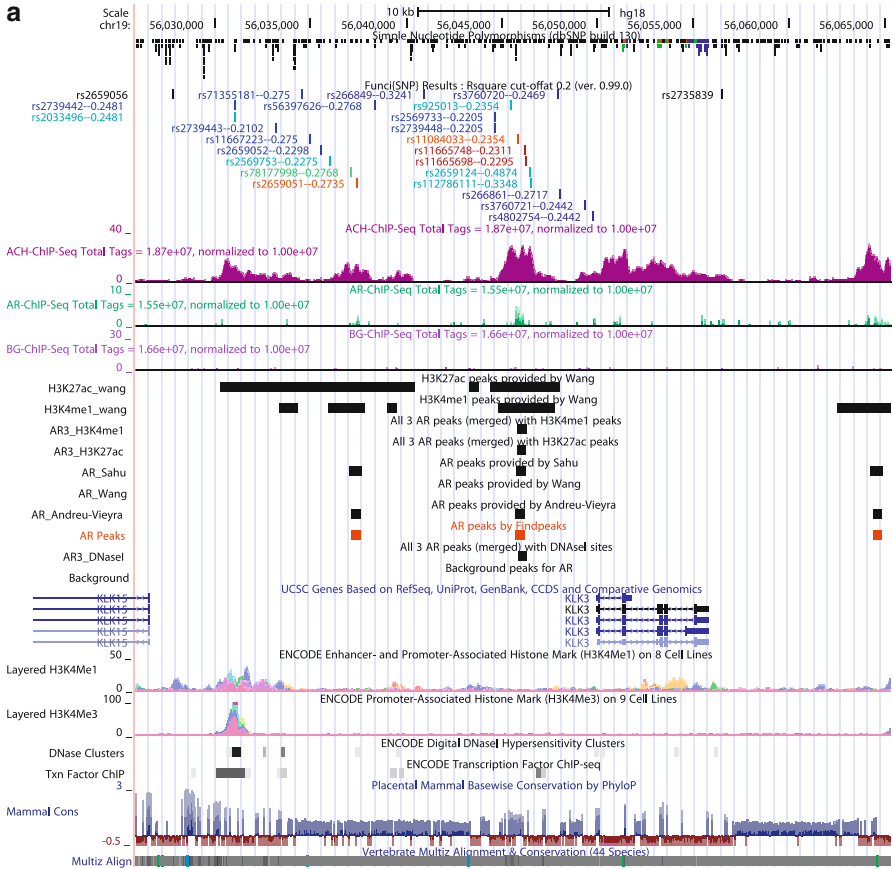
Fig. 5.6 Genomic plots with FunciSNP results and independent validation on eight enhancer elements harboring a candidate functional SNP, performed in three different cell types. (**a**) UCSC genome browser tracks are ordered in the following manner: dbSNP135, FunciSNP result, biofeatures, refseq genes, and known lincRNA. TagSNP is highlighted in the FunciSNP result track, and each candidate SNP is color coded to reflect the number of biofeatures which it overlaps. The color ranges from *blue* (low number of biofeature overlap) to *red* (high number of overlap). Each candidate function SNP is identified by its known rsID and the calculated R^2 value to the known GWAS tagSNP. The results are saved in a UCSC genome session: http://goo.gl/cYCl7. (**b**) Enhancer activities were evaluated using a dual luciferase reporter assay by cloning eight candidate enhancer regions harboring a candidate functional SNP in three different prostate cell types [LNCaP, PC3 (+AR), and PREC (normal primary prostate epithelial cells) (+AR)]. In order to stimulate AR, prostate cells were treated with 10 nM DHT. *Orange–pink* color represents cells treated with DHT and *green–blue* represents cells treated without DHT. Enhancer activities of positive control region (PSA enhancer) ($n=1$) and negative control regions ($n=2$) are presented to the *left*. *Dashed horizontal line* indicates the average enhancer activities of the two negative controls. *Error bar* indicates the standard deviation from the mean of luciferase activity values from four independent experimental replicates
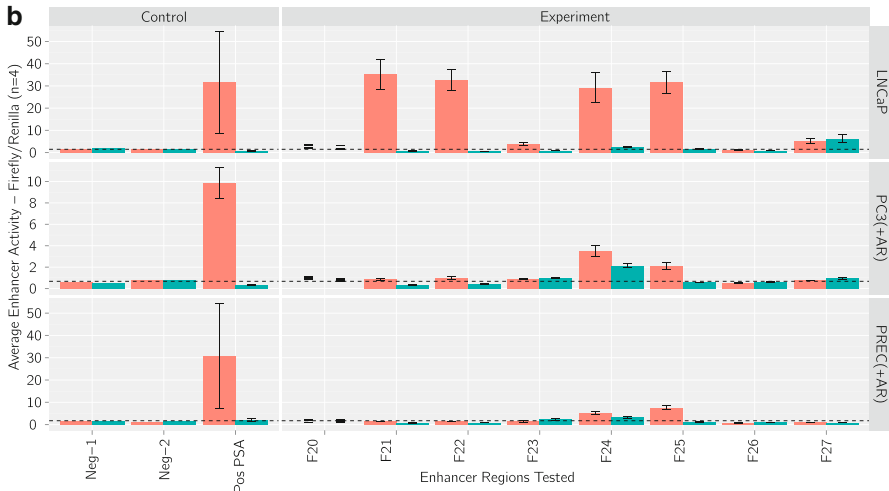
Fig. 5.6 (continued)

Table 5.3 Distribution of total number of 1000 gp overlapping at least one biological features as defined in Table 5.1

	Total SNPs tested	Total number of SNPs coincide with biofeatures ($R.squared>0$)	Total number of correlated SNPs coincide with biofeatures ($R.squared>0.5$)
Number of 1000 genome SNPs	53453	20885	431
Percent (%)	100	39	0.8

Table 5.4 Number of correlated SNPs coincides with biofeatures distributed by each GWAS SNPs for PCa

Column1	1 or more biofeatures	2 or more biofeatures	3 or more biofeatures	4 or more biofeatures	5 or more biofeatures	6 or more biofeatures
rs10086908	32	18	4	0	0	0
rs10090154	18	9	6	1	0	0
rs1016343	7	1	0	0	0	0
rs10187424	18	5	0	0	0	0
rs10486567	4	0	0	0	0	0
rs10875943	3	1	1	1	0	0
rs10896449	3	0	0	0	0	0
rs10896469	1	0	0	0	0	0
rs10934853	2	0	0	0	0	0
rs10936632	3	0	0	0	0	0

(continued)

Table 5.4 (continued)

Column1	1 or more biofeatures	2 or more biofeatures	3 or more biofeatures	4 or more biofeatures	5 or more biofeatures	6 or more biofeatures
rs10993994	19	6	2	1	1	0
rs11649743	2	1	0	0	0	0
rs12155172	1	0	0	0	0	0
rs12418451	7	2	0	0	0	0
rs12500426	12	7	0	0	0	0
rs12621278	5	0	0	0	0	0
rs13252298	17	0	0	0	0	0
rs13254738	1	0	0	0	0	0
rs1327301	4	0	0	0	0	0
rs13385191	2	0	0	0	0	0
rs1447295	18	9	6	1	0	0
rs1456315	1	1	1	1	0	0
rs1465618	1	0	0	0	0	0
rs1512268	6	3	1	0	0	0
rs1571801	1	0	0	0	0	0
rs16901966	27	6	0	0	0	0
rs16901979	27	6	0	0	0	0
rs16902094	8	1	1	1	0	0
rs17021918	5	0	0	0	0	0
rs1859962	9	1	0	0	0	0
rs1983891	2	2	1	0	0	0
rs2292884	37	6	1	0	0	0
rs2660753	4	0	0	0	0	0
rs2928679	4	1	0	0	0	0
rs3123078	24	8	4	2	1	0
rs339331	24	0	0	0	0	0
rs37181	3	1	0	0	0	0
rs3760511	1	0	0	0	0	0
rs4242382	18	9	6	1	0	0
rs4242384	17	9	6	1	0	0
rs4430796	3	0	0	0	0	0
rs445114	4	1	1	1	1	0
rs4962416	2	0	0	0	0	0
rs5919432	9	3	1	0	0	0
rs5945572	7	0	0	0	0	0
rs5945619	7	0	0	0	0	0
rs620861	5	1	1	1	1	0
rs6465657	9	1	0	0	0	0
rs6470494	1	0	0	0	0	0
rs6501455	17	2	0	0	0	0
rs6545977	1	1	1	0	0	0
rs6763931	9	4	1	0	0	0
rs6983267	8	4	0	0	0	0
rs6983561	27	6	0	0	0	0

Table 5.4 (continued)

Column1	1 or more biofeatures	2 or more biofeatures	3 or more biofeatures	4 or more biofeatures	5 or more biofeatures	6 or more biofeatures
rs7000448	8	1	0	0	0	0
rs7127900	24	5	3	0	0	0
rs7210100	2	0	0	0	0	0
rs7214479	18	2	0	0	0	0
rs731236	6	1	0	0	0	0
rs7501939	3	0	0	0	0	0
rs7584330	29	6	1	0	0	0
rs7679673	2	0	0	0	0	0
rs7837688	18	9	6	1	0	0
rs7920517	24	8	4	2	1	0
rs7931342	3	0	0	0	0	0
rs8102476	4	0	0	0	0	0
rs887391	5	1	0	0	0	0
rs902774	5	4	0	0	0	0
rs9284813	4	1	0	0	0	0
rs9311171	5	2	0	0	0	0
rs9623117	3	1	1	1	1	1
rs983085	16	2	0	0	0	0

References

1. Visscher PM et al (2012) Five years of GWAS discovery. Am J Hum Genet 90(1):7–24
2. Hindorff LA et al (2009) Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc Natl Acad Sci USA 106(23):9362–9367
3. Coetzee SG et al (2012) FunciSNP: an R/bioconductor tool integrating functional non-coding data sets with genetic association studies to identify candidate regulatory SNPs. Nucl Acids Res 40:e139
4. Coetzee GA (2012) The usefulness of prostate cancer genome-wide association studies. J Urol 187(1):9–10
5. Coetzee GA et al (2010) A systematic approach to understand the functional consequences of non-protein coding risk regions. Cell Cycle 9(2):47–51
6. Heinlein CA, Chang C (2004) Androgen receptor in prostate cancer. Endocr Rev 25(2):276–308

7. Bluemn EG, Nelson PS (2012) The androgen/androgen receptor axis in prostate cancer. Curr Opin Oncol 24(3):251–257

8. Denmeade SR, Isaacs JT (2002) A history of prostate cancer treatment. Nat Rev Cancer 2(5):389–396

9. Cai C, Balk SP (2011) Intratumoral androgen biosynthesis in prostate cancer pathogenesis and response to therapy. Endocr Relat Cancer 18(5):R175–R182

10. Lonergan PE, Tindall DJ (2011) Androgen receptor signaling in prostate cancer development and progression. J Carcinog 10:20

11. Montgomery RB et al (2008) Maintenance of intratumoral androgens in metastatic prostate cancer: a mechanism for castration-resistant tumor growth. Cancer Res 68(11):4447–4454

12. Mostaghel EA et al (2010) Variability in the androgen response of prostate epithelium to 5alpha-reductase inhibition: implications for prostate cancer chemoprevention. Cancer Res 70(4):1286–1295

13. (2011) Triple-acting drug boosts prostate cancer survival. Cancer Discov 1(7): OF1

14. Tran C et al (2009) Development of a second-generation antiandrogen for treatment of advanced prostate cancer. Science 324(5928):787–790

15. Mukherji D, Pezaro CJ, De-Bono JS (2012) MDV3100 for the treatment of prostate cancer. Expert Opin Investig Drugs 21(2):227–233

16. Grasso CS et al (2012) The mutational landscape of lethal castration-resistant prostate cancer. Nature 487:239–243

17. Chen Y et al (2012) Systematic evaluation of factors influencing ChIP-seq fidelity. Nat Methods 9:609–614

18. Qin B et al (2012) CistromeMap: a knowledgebase and web server for ChIP-Seq and DNase-Seq studies in mouse and human. Bioinformatics 28(10):1411–1412

19. Lupien M et al (2008) FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. Cell 132(6):958–970

20. Andreu-Vieyra C et al (2011) Dynamic nucleosome-depleted regions at androgen receptor enhancers in the absence of ligand in prostate cancer cells. Mol Cell Biol 31(23):4648–4662

21. Wang Q et al (2009) Androgen receptor regulates a distinct transcription program in androgen-independent prostate cancer. Cell 138(2):245–256

22. Wang D et al (2011) Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. Nature 474(7351):390–394

23. Zhang C et al (2011) Definition of a FoxA1 Cistrome that is crucial for G1 to S-phase cell-cycle transit in castration-resistant prostate cancer. Cancer Res 71(21):6738–6748

24. Dryhurst D et al (2012) Histone H2A.Z prepares the prostate specific antigen (PSA) gene for androgen receptor-mediated transcription and is upregulated in a model of prostate cancer progression. Cancer Lett 315(1):38–47

25. Cai C et al (2011) Androgen receptor gene expression in prostate cancer is directly suppressed by the androgen receptor through recruitment of lysine-specific demethylase 1. Cancer Cell 20(4):457–471

26. Sahu B et al (2011) Dual role of FoxA1 in androgen receptor binding to chromatin, androgen signalling and prostate cancer. EMBO J 30(19):3962–3976

27. Taslim C et al (2012) Integrated analysis identifies a class of androgen-responsive genes regulated by short combinatorial long-range mechanism facilitated by CTCF. Nucleic Acids Res 40(11):4754–4764

28. Song L et al (2011) Open chromatin defined by DNaseI and FAIRE identifies regulatory elements that shape cell-type identity. Genome Res 21(10):1757–1767

29. He HH et al (2012) Differential DNase I hypersensitivity reveals factor-dependent chromatin dynamics. Genome Res 22(6):1015–1025

30. Myers RM et al (2011) A user's guide to the encyclopedia of DNA elements (ENCODE). PLoS Biol 9(4):e1001046

31. Rivera A et al (2005) Hypothetical LOC387715 is a second major susceptibility gene for age-related macular degeneration, contributing independently of complement factor H to disease risk. Hum Mol Genet 14(21):3227–3236

32. Hosking FJ, Dobbins SE, Houlston RS (2011) Genome-wide association studies for detecting cancer susceptibility. Br Med Bull 97:27–46
33. Park JH et al (2010) Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. Nat Genet 42(7):570–575
34. Kruglyak L (2008) The road to genome-wide association studies. Nat Rev Genet 9(4):314–318
35. Hirschhorn JN, Daly MJ (2005) Genome-wide association studies for common diseases and complex traits. Nat Rev Genet 6(2):95–108
36. Hardy J, Singleton A (2009) Genomewide association studies and human disease. N Engl J Med 360(17):1759–1768
37. International HapMap Consortium (2003) The International HapMap Project. Nature 426(6968):789–796
38. Pennisi E (2010) Genomics. 1000 Genomes Project gives new map of genetic diversity. Science 330(6004):574–575
39. Consortium TGP (2010) A map of human genome variation from population-scale sequencing. Nature 467(7319):1061–1073
40. Freedman ML et al (2011) Principles for the post-GWAS functional characterization of cancer risk loci. Nat Genet 43(6):513–518
41. Rosenbloom KR et al (2011) ENCODE whole-genome data in the UCSC Genome Browser: update 2012. Nucl Acids Res 40:D912–D917
42. Jia L et al (2009) Functional enhancers at the gene-poor 8q24 cancer-linked locus. PLoS Genet 5(8):e1000597
43. Heinz S et al (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell 38(4):576–589
44. Takata R et al (2010) Genome-wide association study identifies five new susceptibility loci for prostate cancer in the Japanese population. Nat Genet 42(9):751–754
45. Eeles RA et al (2009) Identification of seven new prostate cancer susceptibility loci through a genome-wide association study. Nat Genet 41(10):1116–1121
46. Kote-Jarai Z et al (2011) Seven prostate cancer susceptibility loci identified by a multi-stage genome-wide association study. Nat Genet 43(8):785–791
47. Schumacher FR et al (2011) Genome-wide association study identifies new prostate cancer susceptibility loci. Hum Mol Genet 20(19):3867–3875
48. Murabito JM et al (2007) A genome-wide association study of breast and prostate cancer in the NHLBI's Framingham Heart Study. BMC Med Genet 8(Suppl 1):S6
49. Gudmundsson J et al (2009) Genome-wide association and replication studies identify four variants associated with prostate cancer susceptibility. Nat Genet 41(10):1122–1126
50. Eeles RA et al (2008) Multiple newly identified loci associated with prostate cancer susceptibility. Nat Genet 40(3):316–321
51. Thomas G et al (2008) Multiple loci identified in a genome-wide association study of prostate cancer. Nat Genet 40(3):310–315
52. Zheng SL et al (2007) Association between two unlinked loci at 8q24 and prostate cancer risk among European Americans. J Natl Cancer Inst 99(20):1525–1533
53. Yeager M et al (2007) Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. Nat Genet 39(5):645–649
54. Gudmundsson J et al (2007) Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. Nat Genet 39(5):631–637
55. Duggan D et al (2007) Two genome-wide association studies of aggressive prostate cancer implicate putative prostate tumor suppressor gene DAB2IP. J Natl Cancer Inst 99(24):1836–1844
56. Liu H, Wang B, Han C (2011) Meta-analysis of genome-wide and replication association studies on prostate cancer. Prostate 71(2):209–224
57. Chung CC et al (2011) Fine mapping of a region of chromosome 11q13 reveals multiple independent loci associated with risk of prostate cancer. Hum Mol Genet 20(14):2869–2878

58. Zheng SL et al (2009) Two independent prostate cancer risk-associated Loci at 11q13. Cancer Epidemiol Biomarkers Prev 18(6):1815–1820
59. Bonilla C et al (2011) Prostate cancer susceptibility Loci identified on chromosome 12 in African Americans. PLoS One 6(2):e16044
60. Sun J et al (2008) Evidence for two independent prostate cancer risk-associated loci in the HNF1B gene at 17q12. Nat Genet 40(10):1153–1155
61. Haiman CA et al (2011) Genome-wide association study of prostate cancer in men of African ancestry identifies a susceptibility locus at 17q21. Nat Genet 43(6):570–573
62. Hsu FC et al (2009) A novel prostate cancer susceptibility locus at 19q13. Cancer Res 69(7):2720–2723
63. Sun J et al (2009) Sequence variants at 22q13 are associated with prostate cancer risk. Cancer Res 69(1):10–15