Published in final edited form as:

Prostate. 2012 March ; 72(4): 376–385. doi:10.1002/pros.21439.

Association of prostate cancer risk with SNPs in regions containing androgen receptor binding sites captured by ChIP-on-chip analyses

Yizhen Lu^{1,2,*}, Jielin Sun^{3,*}, Andrew K. Kader^{3,4}, Seong-Tae Kim³, Jin-Woo Kim³, Wennuan Liu³, Jishan Sun³, Daru Lu^{1,2,5}, Junjie Feng³, Yi Zhu³, Tao Jin³, Zheng Zhang³, Latchezar Dimitrov³, James Lowey⁶, Kevin Campbell⁶, Edward Suh⁶, David Duggan⁶, John Carpten⁶, Jeffrey M. Trent^{6,7}, Henrik Gronberg⁸, Siqun L. Zheng³, William B. Isaacs^{9,†}, and Jianfeng Xu^{1,2,3,4,7,†}

¹State Key Laboratory of Genetic Engineering, School of Life Sciences, Fudan University, Shanghai, China

²Fudan-VARI Center for Genetic Epidemiology, School of Life Sciences, Fudan University, Shanghai, China

³Center for Cancer Genomics, Wake Forest University School of Medicine, Winston-Salem, NC

⁴Department of Urology, Wake Forest University School of Medicine, Winston-Salem, NC

⁵MOE Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, Shanghai, China

⁶Translational Genomics Research Institute, Phoenix, AZ

⁷Van Andel Research Institute, Grand Rapids, MI

⁸Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

⁹Department of Urology, Johns Hopkins Medical Institutions, Baltimore, MD

Abstract

Background—Genome-wide association studies (GWAS) have identified approximately three dozen single nucleotide polymorphisms (SNPs) consistently associated with prostate cancer (PCa) risk. Despite the reproducibility of these associations, the molecular mechanism for most of these SNPs has not been well elaborated as most lie within non-coding regions of the genome.

Androgens play a key role in prostate carcinogenesis. Recently, using ChIP-on-chip technology, 22,447 androgen receptor (AR) binding sites have been mapped throughout the genome, greatly expanding the genomic regions potentially involved in androgen-mediated activity.

Methodology/Principal findings—To test the hypothesis that sequence variants in AR binding sites are associated with PCa risk, we performed a systematic evaluation among two existing PCa GWAS cohorts; the Johns Hopkins Hospital and the Cancer Genetic Markers of Susceptibility (CGEMS) study population. We demonstrate that regions containing AR binding sites are significantly enriched for PCa risk-associated SNPs, i.e. more than expected by chance alone. In addition, compared with the entire genome, these newly observed risk-associated SNPs in these regions are significantly more likely to overlap with established PCa risk-associated SNPs

[†]Address for correspondence: Dr. William B. Isaacs, Ph.D., Marburg 115, Johns Hopkins Hospital, 600 N. Wolfe Street, Baltimore, MD 21287, Phone: (410) 955-2518, Fax: (410) 955-0833, wisaacs@jhmi.edu; & Dr. Jianfeng Xu, Center for Cancer Genomics, Medical Center Blvd, Winston-Salem, NC 27157, Phone: (336) 713-7500, Fax: (336) 713-7566, jxu@wfubmc.edu.

*These authors contributed equally

from previous GWAS. These results are consistent with our previous finding from a bioinformatics analysis that one-third of the 33 known PCa risk-associated SNPs discovered by GWAS are located in regions of the genome containing AR binding sites.

Conclusions/Significance—The results to date provide novel statistical evidence suggesting an androgen-mediated mechanism by which some PCa associated SNPs act to influence PCa risk. However, these results are hypothesis generating and ultimately warrant testing through in-depth molecular analyses.

Keywords

AR; prostate cancer; GWAS; pathway association study

Introduction

Prostate cancer (PCa) is the most common non-cutaneous cancer affecting men in the US. Inherited genetic alterations are hypothesized to contribute to the well-characterized familial aggregation of PCa. Since 2007, breakthroughs in genomic technologies have resulted in the discovery of at least 33 PCa risk-associated single nucleotide polymorphisms (SNPs) from genome-wide association studies (GWAS) (1-15). These risk-associated germ-line genetic markers have been consistently replicated in multiple case-control study populations of European descent as well as of African Americans and Asians (16-19). However, because most of these risk-associated SNPs reside in non-coding regions of the genome, the molecular mechanisms by which these SNPs function are uncertain. Recently, we applied a bioinformatics tool to map these PCa risk-associated SNPs to genomic functional annotation databases, including the encyclopedia of DNA elements (ENCODE), 11 genomic regulatory elements databases defined by the University of California Santa Cruz (UCSC) table browser, and androgen receptor (AR)-binding sites (20). We found that PCa risk-associated SNPs were significantly enriched in genomic regions containing AR-binding sites ($P < 0.003$). About one-third of the 33 risk SNP blocks are located within the regions containing AR-binding sites.

Androgens and their receptor (AR) have long been demonstrated to play a critical role in prostate carcinogenesis, and their importance in the development of PCa is further supported by the 23-25% reduction in PCa risk seen in two large clinical trials testing agents which block androgen activation via testosterone to dihydrotestosterone conversion (21-22). AR is a transcription factor which binds to regions of DNA to control the expression of target genes in the presence of androgens. Androgen-response elements (AREs) are the most well-known AR binding sites. The canonical AREs are partial palindromic repeats of 5'-TGTTCT-3' in promoter domains. However, AR binding sites can also be operationally defined as DNA regions that are bound by AR. By combining chromatin immunoprecipitation with tiled oligonucleotide microarrays (ChIP-on-chip), Wang et al performed a genome-wide study and identified 22,447 regions in the genome containing putative AR binding sites, with a median size of 911 base pairs (bp) (ranging from 299 to 5,554 bp) (23). Importantly, most of these regions are not proximal to promoters, and instead map to distal, unannotated *cis*-regulatory elements. These findings have greatly expanded the genomic regions potentially involved in androgen-mediated activity.

In this study, we expanded our research on 33 known PCa risk-associated SNPs and tested the hypothesis that sequence variants in putative AR binding sites across the genome are associated with PCa risk. This was performed by systematically evaluating identified sequence variants in regions containing AR binding sites captured by ChIP-on-chip analyses for their association with PCa risk in two existing PCa GWAS datasets.

Methods

Study populations

Our primary population was from a Johns Hopkins Hospital (JHH) PCa GWAS which included 1,964 PCa cases and 3,172 control subjects. The cases are Caucasian PCa patients who underwent radical prostatectomy for the treatment of PCa at JHH from January 1, 1999, through December 31, 2008 (24). The clinical characteristics of these patients are presented in Supplementary Table 1. The control subjects for this population were an independent group of Caucasian individuals from the Illumina iControlDB (iControls) dataset (25). A second GWAS population was obtained from Stage 1 of the National Cancer Institute Cancer Genetic Markers of Susceptibility (CGEMS) study. It included 1,172 PCa cases and 1,157 control subjects, selected from the Prostate, Lung, Colon and Ovarian (PLCO) Cancer Screening Trial (6,9). The genotype and phenotype data of the study are publicly available and our use of the data was approved by CGEMS.

GWAS genotyping data, imputation and quality control

GWAS of the JHH case population was performed using the Illumina 610K chip (24) and the GWAS of the iControls population (25) was performed using Illumina Hap300 and Hap550 Chips. We imputed all the known SNPs that are catalogued in HapMap Phase II (26) using the IMPUTE computer program (27) with a posterior probability of 0.9 as a threshold to call genotypes. The following quality control criteria were used to filter SNPs: $MAF < 0.01$, $HWE < 0.001$ and call rate < 0.95 .

SNPs in genomic regions containing AR binding sites

The data on the 22,447 putative AR binding sites across the genome discovered by ChIP-on-chip analysis in two PCa cell lines are publically available (23). We performed a search for all known SNPs in these AR binding sites based on the HapMap database (Build 36).

PCa association tests

Allele frequency differences between case patients and control subjects were tested for each SNP in regions of AR binding sites using a chi-square test with 1 degree of freedom. Allelic odds ratio (OR) and 95% confidence intervals (95% CI) were estimated based on a multiplicative model. In order to reduce spurious association results due to potential population stratification in the JHH study, the association tests were performed by adjusting the top five eigenvectors that were estimated using the EIGENSOFT software in the logistic regression analysis (28). A combined allelic test for JHH and CGEMS study populations was performed using the Cochran-Mantel-Haenszel test.

Linkage disequilibrium (LD) block

To identify SNPs that are independently associated with PCa risk, we used the LD block as a unit of association. LD block was defined as a set of SNPs within a genomic region of 1000 kb with pair-wise r^2 value > 0.5 , and was estimated using the CLUMP function of the PLINK software (29).

Reported PCa risk-associated SNPs by GWAS

Based on PCa GWAS reported before December 2009, we selected 33 PCa risk-associated SNPs, each of which exceeded genome-wide significance levels in their initial reports ($P < 10^{-7}$) and all have been replicated in independent study populations (1-15). Considering that many SNPs are in LD with each of these 33 SNPs, we used LD blocks (r^2 value > 0.5) to define the genomic regions of these established PCa risk SNPs based on CEU genotype data from HapMap release#27 (Phase II+PhaseIII). One SNP (rs16902094) was not included in

the study because it was not available in the HapMap release data. Thus there were 32 PCa associated LD blocks. The SNP list and pair-wise r^2 values of the PCa risk-associated SNPs are provided in Supplementary Table 2.

Results

The search for all known SNPs within the regions of 22,447 ChIP-on-chip detected AR binding sites in the HapMap database (Build 36) revealed a total of 18,401 SNPs (Figure 1a-b). These SNPs are listed in Supplementary Table 3. In the JHH study, 12,724 of these 18,401 SNPs were directly genotyped or successfully imputed and passed the quality control criteria outlined in the materials and methods (Figure 1c). Association of PCa risk with each of these 12,724 SNPs was performed using an allelic test. Given that some of these SNPs are in LD, we used the CLUMP function of the PLINK computer program to infer independent LD blocks (Figure 1d). Among the 9,500 independent LD blocks, we observed 4 blocks where at least one SNP was associated with PCa risk at $P < 10^{-5}$ (Figure 1e). This P-value cutoff was chosen to approximate a 5% Type I error after accounting for 9,500 tests using a Bonferroni correction. The four blocks were 8p21 at 3' of *NKX3.1*, 10q11 at 5' of *MSMB*, and regions 1 and 2 of 8q24 (Supplementary Table 4), all of which were known PCa risk-associated SNPs. To assess whether the observation of four significant blocks represented an enrichment of PCa risk-associated SNPs in the regions containing AR binding sites, we performed a simulation analysis under the null hypothesis of no association between SNPs in these 9,500 blocks and PCa risk. Among the 100,000 replicates, we found 10 replicates where 4 or more blocks were significantly associated with PCa risk at $P < 10^{-5}$. Therefore, the excess of observed PCa risk-associated LD blocks (enrichment) was statistically significant, with an empirical P value of 0.0001 (10/100,000) (Figure 1f).

We performed a similar analysis based on GWAS data from an independent study population, CGEMS. Among the 12,369 independent LD blocks containing AR binding sites, we found 2 blocks where at least one SNP was associated with PCa risk at $P < 10^{-5}$ (Figure 1c-f). These two blocks were at 15q21 and the region 1 of 8q24 (Supplementary Table 4), the latter block was also significant in the JHH study at $P < 10^{-5}$. Although not all the specific significant blocks observed in the JHH were replicated, the enrichment of PCa risk-associated SNPs in the regions containing AR binding sites was statistically significant. Based on 100,000 replicates simulated under the null hypothesis of no association between SNPs in these 12,369 blocks and PCa risk, the empirical p value of observing 2 or more blocks that were significantly associated with PCa risk at $P < 10^{-5}$ of a replicate is 0.01. Similar findings were obtained when the two GWAS data were combined (Figure 1c-f).

The enrichment of PCa risk-associated SNPs in the regions containing AR binding sites was also found when less stringent P-value cutoffs (10^{-2} , 10^{-3} , and 10^{-4}) were used to declare significant association, in the JHH, CGEMS, and combined JHH and CGEMS datasets (Supplementary Table 5).

When examining the blocks that are significantly associated with PCa risk, we found that many of them overlapped with LD blocks of known PCa risk-associated SNPs identified from GWAS (i.e., at least one common significant SNP located in both types of LD blocks), and the proportion of overlap was significantly higher than that of remaining significant LD blocks in the genome. For example, among the blocks that were significantly associated with PCa risk at P-value cutoffs of 10^{-2} , 10^{-3} , 10^{-4} , or 10^{-5} in the combined JHH and CGEMS dataset, 4.3%, 26.1%, 71.4%, and 100%, respectively, overlapped with LD blocks containing 32 known PCa risk associated SNPs. In contrast, of the LD blocks that were associated with PCa risk at these P-value cutoffs throughout the genome, only 0.5%, 2.4%, 10.6%, and 42.9%, respectively, overlapped with the blocks of 32 known PCa risk

associated SNPs (Table 1). Based on Fisher's exact tests, the difference between these two types of blocks was statistically significant in most scenarios (at a P-value cutoff of 10^{-2} , 10^{-3} , or 10^{-4}). The difference was not statistically significant at the P-value cutoff of 10^{-5} , likely due to small cell count. These results suggest that higher proportions of observed PCa risk-associated SNPs in the regions containing AR binding sites represent true association.

Discussion

Utilizing GWAS data from two independent PCa association study populations, we obtained two new pieces of statistical evidence supporting an association of sequence variants in regions containing AR binding sites with PCa risk. The first finding was the statistically significant enrichment of PCa risk-associated SNPs in genomic regions containing AR binding sites. The second finding was a significantly higher proportion of overlap between established PCa risk-associated SNPs from previous GWAS, and the newly observed PCa risk-associated SNPs in these AR blocks, compared to other blocks in the genome. These two pieces of evidence, together with our previous result from a bioinformatics analysis demonstrating that 11 of the 33 known PCa risk-associated SNPs reside in regions containing AR binding sites (20), suggest that sequence variants in regions containing AR binding sites are associated with PCa risk. The importance of the findings lies in the provision of a common, potential pathway by which many of the SNPs with strong statistical associations with PCa risk but unknown function act to affect this risk. The findings are plausible because androgen and AR have long been demonstrated to play a critical role in prostate carcinogenesis and because AR binding sites are critical for AR mediated androgen action.

Detailed examination of two exemplified genomic regions where AR binding sites were mapped by Chip-on-chip and SNPs within the regions were highly associated with PCa risk provided further insight on the potential molecular mechanisms of these genetic findings. The first involves the 8p21 genomic region which contains the *NKX3.1* gene that codes for an AR regulated prostate-specific homeodomain-containing transcription factor involved in prostate differentiation and tumor suppression. Sequence variants in the vicinity of *NKX3.1* have been associated with PCa risk by GWAS in several independent study populations (11). In our combined analysis of the JHH and CGEMS dataset, multiple SNPs in a ~50 kb region that includes the *NKX3.1* gene, were strongly associated with PCa ($P < 1.0 \times 10^{-5}$) (Figure 2a, light blue bar). Interestingly, a cluster of five AR binding sites is mapped within or near the ~50 kb PCa risk-associated region. In contrast, no other AR binding sites mapped within the more than 1-Mb interval flanking this region (Figure 2b, short horizontal red and blue lines). The validity of ChIP-on-chip defined AR binding sites at this region was supported by a recently published study where three AREs were identified at the 3' UTR of *NKX3.1* (Figure 2c, three short vertical bars) (30). As shown in Figure 2d, the ARE1 contains 22 bp (sequences in black bold) and is within the 1,470 bp AR binding site mapped by ChIP-on-chip (grey highlighted). Multiple SNPs are listed in this region by the NCBI database. Genotype data were available for two of these SNPs (rs1567669 and rs4872176, red box) in the JHH study. Both SNPs were significantly associated with PCa risk ($P = 10^{-5}$ and 10^{-6} , respectively). However, the closest of these two SNPs (rs4872176) remains 46 bp telomeric to the ARE1. Experiments, beyond the scope of this genetic study, are needed to test whether these sequence variants in the ARE1, and/or variants across the broader region of the AR binding site, may affect the AR mediated action and development of PCa.

The second region is at 8q24, where multiple independent PCa risk associations have been strongly implicated in a 'gene desert' region (11,31-32). In contrast to *NKX3.1*, this region has not been previously associated with AR regulation, and the molecular mechanism by which SNPs at 8q24 affect PCa risk is unclear. Five broadly defined and independent PCa

risk-associated regions at 8q24 have been documented by previous GWAS (Figure 3a, short horizontal blue bars). PCa association at regions 1, 2, 3, and 5 were also found in the combined JHH and CGEMS study (Figure 3b, blue dots). Within a 1.3 Mb span at 8q24 that encompasses the five broadly defined PCa risk-associated regions and two known genes (*MYC* on the telomeric side and *FAM84B* on the centromeric side), regions containing AR binding sites generally concentrated in five broad PCa risk-associated regions (Figure 3c). In particular, three PCa risk-associated regions at 8q24 (region 1, 2, and 4, dotted vertical lines) overlapped with regions containing the AR binding site. The region 3 of 8q24, a locus that is associated with risk to PCa and several non-androgen related cancers, including colorectal and ovarian cancer (33-34), did not overlap with an AR binding site. Of note, a SNP in PCa risk-associated region 3 (rs6983627) was recently reported by another group to have an allelic specific effect on a novel prostate enhancer whose expression mimics that of the *MYC* gene, suggesting that region 3 may affect PCa development through another mechanism (35). We also used Chromatin Immunoprecipitation data and real-time PCR to confirm the AR binding sites within the 8q24 PCa-associated region in the LNCaP cell line (Supplementary methods). All AR binding sites (LNCaP 1-4) detected by ChIP-on-chip at the 8q24 region were confirmed (Supplementary Figure 1). These findings may suggest a novel androgen-dependent mechanism that explains some of the PCa associations at 8q24.

Caution should be exercised when interpreting these results, as this study has several limitations. Given the study design, the statistical evidence presented in this study may represent Type I error (false positive). It is noted that multiple P-value cutoffs (10^{-2} , 10^{-3} , 10^{-4} , and 10^{-5}) were used to test for an enrichment of PCa risk-associated SNPs in genomic regions containing AR binding sites (Figure 1), and to test whether a higher proportion of the blocks containing AR binding sites are associated with established PCa risk SNPs from GWAS (Table 1). The primary purpose for utilizing these different P-value cutoffs was to demonstrate that similar results could be obtained regardless which cutoff was used. However, we recognize that results from different P-value cutoffs were dependent. For example, the number of blocks significantly associated with PCa risk at $P < 10^{-2}$ included the number of blocks significantly associated with PCa risk at $P < 10^{-3}$, 10^{-4} , and 10^{-5} . Therefore, it is difficult to assess the true statistical significance of these findings. Further replication in independent study populations would provide the best assessment of the validity. Second, it is important to emphasize that this study merely provides statistical evidence for a novel hypothesis that some PCa risk-associated sequence variants may mediate their activity through interaction with AR binding sites. These results are hypothesis generating and should ultimately be tested through in-depth molecular analyses. For example, experiments can be designed to test whether implicated SNPs in the AR binding sites affect local AR transcriptional activity, the association of AR or AR co-regulators with chromatin, and androgen signaling. These potential studies, which are beyond the scope of this manuscript, may lead to a better understanding of prostate carcinogenesis. Third, in-depth analyses should be performed to better define AR binding sites. The current AR binding sites were detected using ChIP-on-chip technology in two cell lines. A ChIP-seq approach that combines ChIP and next-generation sequencing could increase the resolution thus better define AR binding sites. When this method was recently applied in other PCa cell lines, it confirmed previously defined AR binding sites and allowed for the identification of novel AR binding sites (36). In addition, the identification of AR binding sites in normal prostate tissues is probably more relevant because the potential influence of SNPs on AR binding sites most likely occurs earlier in normal tissues, prior to the development of PCa. Lastly, additional fine mapping genetic analyses could also be performed by utilizing novel SNPs that were recently discovered in the 1,000 Genomes Project (1KG) (37). However, the false discovery rate of SNPs is relatively high, with an estimated ~10% in the low coverage pilot study of the 1KG study. This limitation may be overcome in the future, when higher

quality catalogues of SNPs, and SNP genotyping chips that include SNPs identified from the 1KG project, become available.

In summary, these findings provide novel statistical evidence for a mechanism by which inherited sequence variants in AR binding sites influence PCa risk. The significance of this novel pathway analysis likely goes beyond AR binding sites and PCa risk. This approach could be applied to sequence variants in other transcription factor binding sites and in other disease settings.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors thank all the study subjects who participated in this study. The authors also thanks for the National Cancer Institute Cancer Genetic Markers of Susceptibility Initiative (CGEMS) for making the data available publicly.

References

1. Amundadottir LT, Sulem P, Gudmundsson J, Helgason A, Baker A, Agnarsson BA, Sigurdsson A, Benediktsdottir KR, Cazier JB, Sainz J, Jakobsdottir M, Kostic J, Magnusdottir DN, Ghosh S, Agnarsson K, Birgisdottir B, Le Roux L, Olafsdottir A, Blondal T, Andresdottir M, Gretarsdottir OS, Bergthorsson JT, Gudbjartsson D, Gylfason A, Thorleifsson G, Manolescu A, Kristjansson K, Geirsson G, Isaksson H, Douglas J, Johansson JE, Bälter K, Wiklund F, Montie JE, Yu X, Suarez BK, Ober C, Cooney KA, Gronberg H, Catalona WJ, Einarsson GV, Barkardottir RB, Gulcher JR, Kong A, Thorsteinsdottir U, Stefansson K. A common variant associated with prostate cancer in European and African populations. *Nat Genet.* 2006; 38:652–658. [PubMed: 16682969]
2. Gudmundsson J, Sulem P, Manolescu A, Amundadottir LT, Gudbjartsson D, Helgason A, Rafnar T, Bergthorsson JT, Agnarsson BA, Baker A, Sigurdsson A, Benediktsdottir KR, Jakobsdottir M, Xu J, Blondal T, Kostic J, Sun J, Ghosh S, Stacey SN, Mouy M, Saemundsdottir J, Backman VM, Kristjansson K, Tres A, Partin AW, Albers-Akkers MT, Godino-Ivan Marcos J, Walsh PC, Swinkels DW, Navarrete S, Isaacs SD, Aben KK, Graif T, Cashy J, Ruiz-Echarri M, Wiley KE, Suarez BK, Witjes JA, Frigge M, Ober C, Jonsson E, Einarsson GV, Mayordomo JI, Kiemeny LA, Isaacs WB, Catalona WJ, Barkardottir RB, Gulcher JR, Thorsteinsdottir U, Kong A, Stefansson K. Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. *Nat Genet.* 2007; 39:631–637. [PubMed: 17401366]
3. Yeager M, Orr N, Hayes RB, Jacobs KB, Kraft P, Wacholder S, Minichiello MJ, Fearnhead P, Yu K, Chatterjee N, Wang Z, Welch R, Staats BJ, Calle EE, Feigelson HS, Thun MJ, Rodriguez C, Albanes D, Virtamo J, Weinstein S, Schumacher FR, Giovannucci E, Willett WC, Cancel-Tassin G, Cussenot O, Valeri A, Andriole GL, Gelmann EP, Tucker M, Gerhard DS, Fraumeni JF Jr, Hoover R, Hunter DJ, Chanock SJ, Thomas G. Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Nat Genet.* 2007; 39:645–649. [PubMed: 17401363]
4. Gudmundsson J, Sulem P, Steinthorsdottir V, Bergthorsson JT, Thorleifsson G, Manolescu A, Rafnar T, Gudbjartsson D, Agnarsson BA, Baker A, Sigurdsson A, Benediktsdottir KR, Jakobsdottir M, Blondal T, Stacey SN, Helgason A, Gunnarsdottir S, Olafsdottir A, Kristinsson KT, Birgisdottir B, Ghosh S, Thorlacius S, Magnusdottir D, Stefansdottir G, Kristjansson K, Bagger Y, Wilensky RL, Reilly MP, Morris AD, Kimber CH, Adeyemo A, Chen Y, Zhou J, So WY, Tong PC, Ng MC, Hansen T, Andersen G, Borch-Johnsen K, Jorgensen T, Tres A, Fuertes F, Ruiz-Echarri M, Asin L, Saez B, van Boven E, Klaver S, Swinkels DW, Aben KK, Graif T, Cashy J, Suarez BK, van Vierssen Trip O, Frigge ML, Ober C, Hofker MH, Wijmenga C, Christiansen C, Rader DJ, Palmer CN, Rotimi C, Chan JC, Pedersen O, Sigurdsson G, Benediktsson R, Jonsson E, Einarsson GV, Mayordomo JI, Catalona WJ, Kiemeny LA, Barkardottir RB, Gulcher JR, Thorsteinsdottir U, Kong A, Stefansson K. Two variants on chromosome 17 confer prostate cancer risk, and the one in TCF2 protects against type 2 diabetes. *Nat Genet.* 2007; 39:977–983. [PubMed: 17603485]

5. Duggan D, Zheng SL, Knowlton M, Benitez D, Dimitrov L, Wiklund F, Robbins C, Isaacs SD, Cheng Y, Li G, Sun J, Chang BL, Marovich L, Wiley KE, Bälter K, Stattin P, Adami HO, Gielzak M, Yan G, Sauvageot J, Liu W, Kim JW, Bleecker ER, Meyers DA, Trock BJ, Partin AW, Walsh PC, Isaacs WB, Grönberg H, Xu J, Carpten JD. Two genome-wide association studies of aggressive prostate cancer implicate putative prostate tumor suppressor gene DAB2IP. *J Natl Cancer Inst*. 2007; 99:1836–1844. [PubMed: 18073375]
6. Thomas G, Jacobs KB, Yeager M, Kraft P, Wacholder S, Orr N, Yu K, Chatterjee N, Welch R, Hutchinson A, Crenshaw A, Cancel-Tassin G, Staats BJ, Wang Z, Gonzalez-Bosquet J, Fang J, Deng X, Berndt SI, Calle EE, Feigelson HS, Thun MJ, Rodriguez C, Albanes D, Virtamo J, Weinstein S, Schumacher FR, Giovannucci E, Willett WC, Cussenot O, Valeri A, Andriole GL, Crawford ED, Tucker M, Gerhard DS, Fraumeni JF Jr, Hoover R, Hayes RB, Hunter DJ, Chanock SJ. Multiple loci identified in a genome-wide association study of prostate cancer. *Nat Genet*. 2008; 40:310–315. [PubMed: 18264096]
7. Gudmundsson J, Sulem P, Rafnar T, Bergthorsson JT, Manolescu A, Gudbjartsson D, Agnarsson BA, Sigurdsson A, Benediktsdottir KR, Blondal T, Jakobsdottir M, Stacey SN, Kostic J, Kristinsson KT, Birgisdottir B, Ghosh S, Magnusdottir DN, Thorlacius S, Thorleifsson G, Zheng SL, Sun J, Chang BL, Elmore JB, Breyer JP, McReynolds KM, Bradley KM, Yaspan BL, Wiklund F, Stattin P, Lindström S, Adami HO, McDonnell SK, Schaid DJ, Cunningham JM, Wang L, Cerhan JR, St Sauver JL, Isaacs SD, Wiley KE, Partin AW, Walsh PC, Polo S, Ruiz-Echarri M, Navarrete S, Fuertes F, Saez B, Godino J, Weijerman PC, Swinkels DW, Aben KK, Witjes JA, Suarez BK, Helfand BT, Frigge ML, Kristjansson K, Ober C, Jonsson E, Einarsson GV, Xu J, Gronberg H, Smith JR, Thibodeau SN, Isaacs WB, Catalona WJ, Mayordomo JI, Kiemeny LA, Barkardottir RB, Gulcher JR, Thorsteinsdottir U, Kong A, Stefansson K. Common sequence variants on 2p15 and Xp11.22 confer susceptibility to prostate cancer. *Nat Genet*. 2008; 40:281–283. [PubMed: 18264098]
8. Eeles RA, Kote-Jarai Z, Giles GG, Olama AA, Guy M, Jugurnauth SK, Mulholland S, Leongamornlert DA, Edwards SM, Morrison J, Field HI, Southey MC, Severi G, Donovan JL, Hamdy FC, Dearnaley DP, Muir KR, Smith C, Bagnato M, Arden-Jones AT, Hall AL, O'Brien LT, Gehr-Swain BN, Wilkinson RA, Cox A, Lewis S, Brown PM, Jhavar SG, Tymrakiewicz M, Lophatananon A, Bryant SL, UK Genetic Prostate Cancer Study Collaborators; British Association of Urological Surgeons' Section of Oncology; UK ProtecT Study Collaborators. Horwich A, Huddart RA, Khoo VS, Parker CC, Woodhouse CJ, Thompson A, Christmas T, Ogden C, Fisher C, Jamieson C, Cooper CS, English DR, Hopper JL, Neal DE, Easton DF. Multiple newly identified loci associated with prostate cancer susceptibility. *Nat Genet*. 2008; 40:316–321. [PubMed: 18264097]
9. Yeager M, Chatterjee N, Ciampa J, Jacobs KB, Gonzalez-Bosquet J, Hayes RB, Kraft P, Wacholder S, Orr N, Berndt S, Yu K, Hutchinson A, Wang Z, Amundadottir L, Feigelson HS, Thun MJ, Diver WR, Albanes D, Virtamo J, Weinstein S, Schumacher FR, Cancel-Tassin G, Cussenot O, Valeri A, Andriole GL, Crawford ED, Haiman CA, Henderson B, Kolonel L, Le Marchand L, Siddiq A, Riboli E, Key TJ, Kaaks R, Isaacs W, Isaacs S, Wiley KE, Gronberg H, Wiklund F, Stattin P, Xu J, Zheng SL, Sun J, Vatten LJ, Hveem K, Kumle M, Tucker M, Gerhard DS, Hoover RN, Fraumeni JF Jr, Hunter DJ, Thomas G, Chanock SJ. Identification of a new prostate cancer susceptibility locus on chromosome 8q24. *Nat Genet*. 2009; 41:1055–1057. [PubMed: 19767755]
10. Gudmundsson J, Sulem P, Gudbjartsson DF, Blondal T, Gylfason A, Agnarsson BA, Benediktsdottir KR, Magnusdottir DN, Orlygsdottir G, Jakobsdottir M, Stacey SN, Sigurdsson A, Wahlfors T, Tammela T, Breyer JP, McReynolds KM, Bradley KM, Saez B, Godino J, Navarrete S, Fuertes F, Murillo L, Polo E, Aben KK, van Oort IM, Suarez BK, Helfand BT, Kan D, Zanon C, Frigge ML, Kristjansson K, Gulcher JR, Einarsson GV, Jonsson E, Catalona WJ, Mayordomo JI, Kiemeny LA, Smith JR, Schleutker J, Barkardottir RB, Kong A, Thorsteinsdottir U, Rafnar T, Stefansson K. Genome-wide association and replication studies identify four variants associated with prostate cancer susceptibility. *Nat Genet*. 2009; 41:1122–1126. [PubMed: 19767754]
11. Eeles RA, Kote-Jarai Z, Al Olama AA, Giles GG, Guy M, Severi G, Muir K, Hopper JL, Henderson BE, Haiman CA, Schleutker J, Hamdy FC, Neal DE, Donovan JL, Stanford JL, Ostrander EA, Ingles SA, John EM, Thibodeau SN, Schaid D, Park JY, Spurdle A, Clements J, Dickinson JL, Maier C, Vogel W, Dörk T, Rebbeck TR, Cooney KA, Cannon-Albright L, Chappuis PO, Hutter P, Zeegers M, Kaneva R, Zhang HW, Lu YJ, Foulkes WD, English DR,

- Leongamornlert DA, Tymrakiewicz M, Morrison J, Arden-Jones AT, Hall AL, O'Brien LT, Wilkinson RA, Saunders EJ, Page EC, Sawyer EJ, Edwards SM, Dearnaley DP, Horwich A, Huddart RA, Khoo VS, Parker CC, Van As N, Woodhouse CJ, Thompson A, Christmas T, Ogden C, Cooper CS, Southey MC, Lophatananon A, Liu JF, Kolonel LN, Le Marchand L, Wahlfors T, Tammela TL, Auvinen A, Lewis SJ, Cox A, FitzGerald LM, Koopmeiners JS, Karyadi DM, Kwon EM, Stern MC, Corral R, Joshi AD, Shahabi A, McDonnell SK, Sellers TA, Pow-Sang J, Chambers S, Aitken J, Gardiner RA, Batra J, Kedda MA, Lose F, Polanowski A, Patterson B, Serth J, Meyer A, Luedeke M, Stefflova K, Ray AM, Lange EM, Farnham J, Khan H, Slavov C, Mitkova A, Cao G, UK Genetic Prostate Cancer Study Collaborators/British Association of Urological Surgeons' Section of Oncology; UK ProtecT Study Collaborators; PRACTICAL Consortium. Easton DF. Identification of seven new prostate cancer susceptibility loci through a genome-wide association study. *Nat Genet.* 2009; 41:1116–1121. [PubMed: 19767753]
12. Duggan D, Zheng SL, Knowlton M, Benitez D, Dimitrov L, Wiklund F, Robbins C, Isaacs SD, Cheng Y, Li G, Sun J, Chang BL, Marovich L, Wiley KE, Bälter K, Stattin P, Adami HO, Gielzak M, Yan G, Sauvageot J, Liu W, Kim JW, Bleecker ER, Meyers DA, Trock BJ, Partin AW, Walsh PC, Isaacs WB, Grönberg H, Xu J, Carpten JD. Two genome-wide association studies of aggressive prostate cancer implicate putative prostate tumor suppressor gene DAB2IP. *J Natl Cancer Inst.* 2007; 99:1836–1844. [PubMed: 18073375]
 13. Sun J, Zheng SL, Wiklund F, Isaacs SD, Purcell LD, Gao Z, Hsu FC, Kim ST, Liu W, Zhu Y, Stattin P, Adami HO, Wiley KE, Dimitrov L, Sun J, Li T, Turner AR, Adams TS, Adolfsson J, Johansson JE, Lowey J, Trock BJ, Partin AW, Walsh PC, Trent JM, Duggan D, Carpten J, Chang BL, Grönberg H, Isaacs WB, Xu J. Evidence for two independent prostate cancer risk-associated loci in the HNF1B gene at 17q12. *Nat Genet.* 2008; 40:1153–1155. [PubMed: 18758462]
 14. Hsu FC, Sun J, Wiklund F, Isaacs SD, Wiley KE, Purcell LD, Gao Z, Stattin P, Zhu Y, Kim ST, Zhang Z, Liu W, Chang BL, Walsh PC, Duggan D, Carpten JD, Isaacs WB, Grönberg H, Xu J, Zheng SL. A novel prostate cancer susceptibility locus at 19q13. *Cancer Res.* 2009; 69:2720–2723. [PubMed: 19318570]
 15. Sun J, Zheng SL, Wiklund F, Isaacs SD, Li G, Wiley KE, Kim ST, Zhu Y, Zhang Z, Hsu FC, Turner AR, Stattin P, Liu W, Kim JW, Duggan D, Carpten J, Isaacs W, Grönberg H, Xu J, Chang BL. Sequence variants at 22q13 are associated with prostate cancer risk. *Cancer Res.* 2009; 69:10–15. [PubMed: 19117981]
 16. Kote-Jarai Z, Easton DF, Stanford JL, Ostrander EA, Schleutker J, Ingles SA, Schaid D, Thibodeau S, Dörk T, Neal D, Donovan J, Hamdy F, Cox A, Maier C, Vogel W, Guy M, Muir K, Lophatananon A, Kedda MA, Spurdle A, Steginga S, John EM, Giles G, Hopper J, Chappuis PO, Hutter P, Foulkes WD, Hamel N, Salinas CA, Koopmeiners JS, Karyadi DM, Johanneson B, Wahlfors T, Tammela TL, Stern MC, Corral R, McDonnell SK, Schürmann P, Meyer A, Kuefer R, Leongamornlert DA, Tymrakiewicz M, Liu JF, O'Mara T, Gardiner RA, Aitken J, Joshi AD, Severi G, English DR, Southey M, Edwards SM, Al Olama AA, PRACTICAL Consortium. Eeles RA. Multiple novel prostate cancer predisposition loci confirmed by an international study: the PRACTICAL Consortium. *Cancer Epidemiol Biomarkers Prev.* 2008; 17:2052–2061. [PubMed: 18708398]
 17. Chang BL, Spangler E, Gallagher S, Haiman CA, Henderson B, Isaacs W, Benford ML, Kidd LR, Cooney K, Strom S, Ingles SA, Stern MC, Corral R, Joshi AD, Xu J, Giri VN, Rybicki B, Neslund-Dudas C, Kibel AS, Thompson IM, Leach RJ, Ostrander EA, Stanford JL, Witte J, Casey G, Eeles R, Hsing AW, Chanock S, Hu JJ, John EM, Park J, Stefflova K, Zeigler-Johnson C, Rebbeck TR. Validation of genome-wide prostate cancer associations in men of African descent. *Cancer Epidemiol Biomarkers Prev.* 2011; 20:23–32. [PubMed: 21071540]
 18. Zheng SL, Hsing AW, Sun J, Chu LW, Yu K, Li G, Gao Z, Kim ST, Isaacs WB, Shen MC, Gao YT, Hoover RN, Xu J. Association of 17 prostate cancer susceptibility loci with prostate cancer risk in Chinese men. *Prostate.* 2010; 70:425–432. [PubMed: 19866473]
 19. Yamada H, Penney KL, Takahashi H, Katoh T, Yamano Y, Yamakado M, Kimura T, Kuruma H, Kamata Y, Egawa S, Freedman ML. Replication of prostate cancer risk loci in a Japanese case-control association study. *J Natl Cancer Inst.* 2009; 101:1330–1336. [PubMed: 19726753]
 20. Lu Y, Zhang Z, Yu H, Zheng SL, Isaacs WB, Xu J, Sun J. Functional annotation of risk loci identified through genome-wide association studies for prostate cancer. *Prostate.* 2011; 71:955–963. [PubMed: 21541972]

21. Thompson IM, Goodman PJ, Tangen CM, Lucia MS, Miller GJ, Ford LG, Lieber MM, Cespedes RD, Atkins JN, Lippman SM, Carlin SM, Ryan A, Szczepanek CM, Crowley JJ, Coltman CA Jr. The influence of finasteride on the development of prostate cancer. *N Engl J Med*. 2003; 349:215–224. [PubMed: 12824459]
22. Andriole GA, Bostwick D, Brawley OW. The influence of dutasteride on the risk of biopsy-detectable prostate cancer: Outcomes of the REDuction by DUtasteride of Prostate Cancer Events (REDUCE) study. *N Engl J Med*. 2010 In press.
23. Wang Q, Li W, Zhang Y, Yuan X, Xu K, Yu J, Chen Z, Beroukhi R, Wang H, Lupien M, Wu T, Regan MM, Meyer CA, Carroll JS, Manrai AK, Jänne OA, Balk SP, Mehra R, Han B, Chinnaiyan AM, Rubin MA, True L, Fiorentino M, Fiore C, Loda M, Kantoff PW, Liu XS, Brown M. Androgen receptor regulates a distinct transcription program in androgen-independent prostate cancer. *Cell*. 2009; 138:245–256. [PubMed: 19632176]
24. Xu J, Zheng SL, Isaacs SD, Wiley KE, Wiklund F, Sun J, Kader AK, Li G, Purcell LD, Kim ST, Hsu FC, Stattin P, Hugosson J, Adolfsson J, Walsh PC, Trent JM, Duggan D, Carpten J, Grönberg H, Isaacs WB. Inherited genetic variant predisposes to aggressive but not indolent prostate cancer. *Proc Natl Acad Sci U S A*. 2010; 107:2136–2140. [PubMed: 20080650]
25. San Diego: Illumina, Inc.; c2010. Illumina iContraolDB [website on the Internet]. Available from: <http://www.illumina.com/science/icontroldb.ilmn>
26. International HapMap Project [database on the Internet]. Available from: <http://hapmap.ncbi.nlm.nih.gov/>
27. Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet*. 2007; 39:906–913. [PubMed: 17572673]
28. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*. 2006; 38:904–909. [PubMed: 16862161]
29. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, Sham PC. PLINK: a toolset for whole-genome association and population-based linkage analysis. *Am J Hum Genet*. 2007; 81:559–575. [PubMed: 17701901]
30. Thomas MA, Preece DM, Bentel JM. Androgen regulation of the prostatic tumour suppressor NKX3.1 is mediated by its 3' untranslated region. *Biochem J*. 2010; 425:575–583. [PubMed: 19886863]
31. Al Olama AA, Kote-Jarai Z, Giles GG, Guy M, Morrison J, Severi G, Leongamornlert DA, Tymrakiewicz M, Jhavar S, Saunders E, Hopper JL, Southey MC, Muir KR, English DR, Dearnaley DP, Ardern-Jones AT, Hall AL, O'Brien LT, Wilkinson RA, Sawyer E, Lophatananon A, UK Genetic Prostate Cancer Study Collaborators/British Association of Urological Surgeons' Section of Oncology; UK Prostate testing for cancer and Treatment study (ProtecT Study) Collaborators. Horwich A, Huddart RA, Khoo VS, Parker CC, Woodhouse CJ, Thompson A, Christmas T, Ogden C, Cooper C, Donovan JL, Hamdy FC, Neal DE, Eeles RA, Easton DF. Multiple loci on 8q24 associated with prostate cancer susceptibility. *Nat Genet*. 2009; 41:1058–1060. [PubMed: 19767752]
32. Yeager M, Chatterjee N, Ciampa J, Jacobs KB, Gonzalez-Bosquet J, Hayes RB, Kraft P, Wacholder S, Orr N, Berndt S, Yu K, Hutchinson A, Wang Z, Amundadottir L, Feigelson HS, Thun MJ, Diver WR, Albanes D, Virtamo J, Weinstein S, Schumacher FR, Cancel-Tassin G, Cussenot O, Valeri A, Andriole GL, Crawford ED, Haiman CA, Henderson B, Kolonel L, Le Marchand L, Siddiq A, Riboli E, Key TJ, Kaaks R, Isaacs W, Isaacs S, Wiley KE, Gronberg H, Wiklund F, Stattin P, Xu J, Zheng SL, Sun J, Vatten LJ, Hveem K, Kumle M, Tucker M, Gerhard DS, Hoover RN, Fraumeni JF Jr, Hunter DJ, Thomas G, Chanock SJ. Identification of a new prostate cancer susceptibility locus on chromosome 8q24. *Nat Genet*. 2009; 41:1055–1057. [PubMed: 19767755]
33. Tomlinson I, Webb E, Carvajal-Carmona L, Broderick P, Kemp Z, Spain S, Penegar S, Chandler I, Gorman M, Wood W, Barclay E, Lubbe S, Martin L, Sellick G, Jaeger E, Hubner R, Wild R, Rowan A, Fielding S, Howarth K, CORGI Consortium. Silver A, Atkin W, Muir K, Logan R, Kerr D, Johnstone E, Sieber O, Gray R, Thomas H, Peto J, Cazier JB, Houlston R. A genome-wide

- association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. *Nat Genet.* 2007; 39:984–988. [PubMed: 17618284]
34. Ghoussaini M, Song H, Koessler T, Al Olama AA, Kote-Jarai Z, Driver KE, Pooley KA, Ramus SJ, Kjaer SK, Hogdall E, DiCioccio RA, Whittemore AS, Gayther SA, Giles GG, Guy M, Edwards SM, Morrison J, Donovan JL, Hamdy FC, Dearnaley DP, Ardern-Jones AT, Hall AL, O'Brien LT, Gehr-Swain BN, Wilkinson RA, Brown PM, Hopper JL, Neal DE, Pharoah PD, Ponder BA, Eeles RA, Easton DF, Dunning AM. UK Genetic Prostate Cancer Study Collaborators/British Association of Urological Surgeons' Section of Oncology; UK ProtecT Study Collaborators. Multiple loci with different cancer specificities within the 8q24 gene desert. *J Natl Cancer Inst.* 2008; 100:962–966. [PubMed: 18577746]
 35. Wasserman NF, Aneas I, Nobrega MA. An 8q24 gene desert variant associated with prostate cancer risk confers differential in vivo activity to a MYC enhancer. *Genome Res.* 2010; 20:1191–1197. [PubMed: 20627891]
 36. Yu J, Yu J, Mani RS, Cao Q, Brenner CJ, Cao X, Wang X, Wu L, Li J, Hu M, Gong Y, Cheng H, Laxman B, Vellaichamy A, Shankar S, Li Y, Dhanasekaran SM, Morey R, Barrette T, Lonigro RJ, Tomlins SA, Varambally S, Qin ZS, Chinnaiyan AM. An integrated network of androgen receptor, polycomb, and TMPRSS2-ERG gene fusions in prostate cancer progression. *Cancer Cell.* 2010; 17:443–454. [PubMed: 20478527]
 37. Durbin RM, Abecasis GR, Altshuler DL, Auton A, Brooks LD, Durbin RM, Gibbs RA, Hurles ME, McVean GA. 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature.* 2010; 467:1061–1073. [PubMed: 20981092]

Figure 1.

Flow chart of study design and key findings: a) number of AR binding sites in the genome defined by chromatin immunoprecipitation with tiled oligonucleotide microarrays (ChIP-on-chip) reported by Wang et al. (2009), b) number of known SNPs in the regions containing AR binding sites, c) number of SNPs that were directly genotyped or imputed with a call rate $\geq 95\%$ in each study in the JHH, CGEMS, and combined JHH and CGEMS dataset, d) number of independent LD blocks, e) number of LD blocks where at least one SNP in the block was statistically associated with PCa risk (allelic test) in each study, and f) results from simulation analysis using a bootstrapping method.

a. PCa association in JHH & CGEMS, LD blocks, Entrez genes

b. AR binding sites LNCaP ABL

c. Androgen response elements (ARE)

8p21, kb

d. Detailed view of ARE1 and SNPs

23,594,895	CACGGAGACC	CAAGTGAAGA	TATGGTTCCA	GAACAGACGC	TATTAAGACTA	rs35350906
23,594,845	AGCGAAGCA	GCTCTCCTCG	GAGCTGGGAG	ACTTGGAGAA	GCACTCCTCT	
23,594,795	TGCGGGGCC	TGAAAGAGGA	GGCCTTCTCC	CGGGCCTCCC	TGGTCTCCGT	rs35675040
23,594,745	GTATAACAGC	TATCTTACT	ACCCATACCT	GTAAGTGGTG	GGCAGCTGGA	
23,594,695	GCCAGCTTTT	TTGGTAATgc	cagctcaggt	gacaaccatt	atgatcaaaa	
23,594,645	actgccttcc	ccagggtgtc	tctatgaaaa	gcacaagggg	ccaaggtcag	
23,594,595	ggagcaagag	gtgtgcacac	caaagctatt	ggagatttgc	gtggaatctc	
23,594,545	cagattcttc	actggtgaga	caatgaaaca	acagagacag	tgaaagtctt	
23,594,495	aatacctaag	tcattctcc	agtgcatact	gtaggtcatt	ttttttgctt	rs1567669
23,594,445	ctggctacct	gtttgaagg	gagagaggga	aaatcaagtg	gtattttcca	
23,594,395	gcactttgta	tgattttgga	tgagtgttac	acccaaggat	tctgtcttgc	rs4872177
23,594,345	aactccatcc	tcctgtgtca	ctgaatatca	actctgaaag	agcaaaccta	
23,594,295	acaggagaaa	ggacaaccag	gatgaggatg	tcaccaactg	aattaaactt	
23,594,245	aagtcacaga	gcctcctgtt	ggccttgga	tatggccaag	gctctctctg	
23,594,195	tcctgttaaa	agagaggggc	aaatagagag	tctccaagag	aacgccctca	
23,594,145	tgctcagcac	atatttgcac	gggaggggga	gatgggtggg	aggagatgaa	
23,594,095	aatacagct	tttcttattc	ctttttattc	cttttaaaat	ggtatgccaa	
23,594,045	cttaagtatt	tacagggtgg	cccaaataga	acaagatgca	ctcgtctgtg	ARE1
23,593,995	ttttaagaca	agctgtataa	acagaactcc	actgcaagag	ggggggccgg	rs28656980
23,593,945	gccaggagaa	tctccgcttg	tccaagacag	gggcctaagg	agggtctcca	rs4872176
23,593,895	cactgctgct	aggggctgtt	gcattttttt	attagtagaa	agtggaaaag	
23,593,845	cctcttctca	acttttttcc	cttgggcttg	agaatttaga	atcagaagtt	rs55982585
23,593,795	tcttgagatt	ttcaggctat	catatatact	gtatcctgaa	aggcaacata	
23,593,745	attcttctct	ccctctcttt	aaaattttgt	gttctctttt	gcagcaatta	
23,593,695	ctcactaaag	ggcttcattt	tagtccagat	ttttagctct	gctgcacctc	
23,593,645	acttatgcct	cgcttattta	gcocagagac	tgggtttttt	tttttttttt	rs28391595
23,593,595	tttttttttt	ttcgtctctc	ccaaagcttt	atctgtcttg	actttttaaa	
23,593,545	aaagtgtggg	ggcagattct	gaattggcta	aaagacatgc	atttttaaaa	
23,593,495	ctagcaactc	ttatttcttt	ccttttaaaa	tacatagcat	taaatcccaa	
23,593,445	atcctattta	aagacctgac	agcttgagaa	ggctcactac	gcatttatag	

Figure 2.

PCa risk-associated SNPs and AR binding sites near *NKX3.1* at 8p21: a) association of SNPs at 8p21 with PCa risk in the combined JHH/CGEMS analysis, b) AR binding sites at the region defined by ChIP-on-chip method (Wang et al. 2009), c) AREs of *NKX3.1* reported by Thomas et al (2010), d) DNA sequences of the most centromeric AR binding site at the region, *NKX3.1* ARE1, and known SNPs.

Figure 3.

PCa risk-associated SNPs and AR binding sites at 8q24: a) association of SNPs at 8q24 with PCa risk in a large published study (Olama et al 2009), b) association of SNPs at 8q24 with PCa risk in the combined JHH/CGEMS analysis, c) AR binding sites at the region defined by ChIP-on-chip methods (Wang et al. 2009), d) known genes at the region. The dashed vertical lines indicate SNPs that are located in the AR binding region.

Table 1
Proportion of PCa risk-associated blocks in the regions containing AR binding sites or in the genome that overlap with known PCa risk-associated SNPs

P-value cutoff for declaring significant association	Number and (%) of significant blocks that overlap with known PCa risk-associated SNPs in regions of		
	AR binding sites	genome	Fisher's exact test
$<10^{-2}$	6/139 (4.3%)	29/5847 (0.5%)	0.0001
$<10^{-3}$	6/23 (26.1%)	20/821 (2.4%)	0.00003
$<10^{-4}$	5/7 (71.4%)	14/132 (10.6%)	0.0005
$<10^{-5}$	3/3 (100%)	9/21 (42.9%)	0.22

The results were based on the combined GWAS dataset of JHH and CGEMS

JHH: Johns Hopkins Hospital

CGEMS: National Cancer Institute Cancer Genetic Markers of Susceptibility

Known prostate cancer (PCa) risk-associated SNPs were identified from previous GWAS (1-15).