Single and Multigenic Analysis of the Association between Variants in 12 Steroid Hormone Metabolism Genes and Risk of Prostate Cancer

Joke Beuten,[1,2] Jonathan A.L. Gelfond,[3] Jennifer L. Franke,[1] Korri S. Weldon,[1] AnaLisa C. Crandall,[1] Teresa L. Johnson-Pais,[2] Ian M. Thompson,[4] and Robin J. Leach[1,2,4]

Departments of [1]Cellular and Structural Biology, [2]Pediatrics, [3]Epidemiology and Biostatistics, and [4]Urology, The University of Texas Health Science Center, San Antonio, Texas

Abstract

To estimate the prostate cancer risk conferred by individual single nucleotide polymorphisms (SNPs), SNP-SNP interactions, and/or cumulative SNP effects, we evaluated the association between prostate cancer risk and the genetic variants of 12 key genes within the steroid hormone pathway (*CYP17, HSD17B3, ESR1, SRD5A2, HSD3B1, HSD3B2, CYP19, CYP1A1, CYP1B1, CYP3A4, CYP27B1*, and *CYP24A1*). A total of 116 tagged SNPs covering the group of genes were analyzed in 2,452 samples (886 cases and 1,566 controls) in three ethnic/racial groups. Several SNPs within *CYP19* were significantly associated with prostate cancer in all three ethnicities (*P* = 0.001-0.009). Genetic variants within *HSD3B2* and *CYP24A1* conferred increased risk of prostate cancer in non-Hispanic or Hispanic Caucasians. A significant gene-dosage effect for increasing numbers of potential high-risk genotypes was found in non-Hispanic and Hispanic Caucasians. Higher-order interactions showed a seven-SNP interaction involving *HSD17B3, CYP19*, and *CYP24A1* in Hispanic Caucasians (*P* = 0.001). In African Americans, a 10-locus model, with SNPs located within *SRD5A2, HSD17B3, CYP17, CYP27B1, CYP19*, and *CYP24A1*, showed a significant interaction (*P* = 0.014). In non-Hispanic Caucasians, an interaction of four SNPs in *HSD3B2, HSD17B3*, and *CYP19* was found (*P* < 0.001). These data are consistent with a polygenic model of prostate cancer, indicating that multiple interacting genes of the steroid hormone pathway confer increased risk of prostate cancer. (Cancer Epidemiol Biomarkers Prev 2009;18(6):1869–80)

Introduction

The underlying etiology of prostate cancer is poorly understood, with genetic predisposition and environmental factors likely contributing to risk (1). Heritable risk factors can involve highly penetrant susceptibility genes with low frequencies and/or low-penetrant genes with higher frequencies in the population. The majority of prostate cancer cases likely involve more common, low-penetrant to moderate-penetrant alleles in genes that are components of functional prostatic pathways, such as androgen hormones and their metabolizing enzymes (2-4).

Evidence exists that the steroid hormone pathway and genes involved in the metabolism of estrogens and androgens affect the risk of prostate cancer (4). Androgens are essential for the development, growth, and secretory activities of the prostate, whereas estrogens modulate these effects. Several of the cytochrome p450 enzymes participate in the synthesis of testosterone from cholesterol (ref. 5; Fig. 1). The *CYP17* gene (10q24.3, MIM 609300) encodes steroid 17α-hydroxylase and is a member of the cytochrome P450 super family. *CYP17* mediates both 17α-hydroxylase and 17,20-lyase activity, and is involved in the biosynthesis of both 17α-hydroxylated glucocorticoids and sex steroids (6-8). The final step in the synthesis of testosterone is the conversion of androstenedione to testosterone by 17β-hydroxysteroid dehydrogenase type 3 (*HSD17B3*, 9q22, MIM 605573). Testosterone, along with estrone and estradiol, interact with the estrogen receptor *ESR1* (6q25.1, MIM 133430), leading to the activation of the enzyme steroid 5α-reductase type 2 (*SRD5A2*, 2p23, MIM 607306). Testosterone is irreversibly metabolized to dihydrotestosterone by *SRD5A2*. Normal prostate growth and development requires dihydrotestosterone and a functioning of *SRD5A2* (4). Dihydrotestosterone stimulates the transcription of several genes with androgen-responsive elements in their promoters and is inactivated in the prostate by hydroxy-δ5-steroid dehydrogenase, 3β and steroid δ-isomerase 1 (*HSD3B1*, 1p13.1, MIM 109715), hydroxy-δ5-steroid dehydrogenase, and 3β and steroid δ-isomerase 2 (*HSD3B2*, 1p13.1, MIM 201810). In the prostate, androstenedione and testosterone can also be converted by aromatase (cytochrome p450, family 19, subfamily A, polypeptide 1; *CYP19*, 15q21.1, MIM 107910) into estrone and estradiol, which are then metabolized by genes from the cytochrome P450 family, including *CYP1A1* (15q22-q24, MIM 108330), *CYP1B1* (2p21, MIM 601771), and *CYP3A4*

Received 1/28/09; revised 3/9/09; accepted 4/9/09; published online 6/8/09.

Grant support: National Cancer Institute Early Detection Research Network grant 5U01CA086402, American Cancer Society grant TURSG-03-152-01-CCE entitled "The Role of Genetic Variation in Prostate Cancer among Hispanics and Blacks," and National Cancer Institute Cancer Support Grant P30CA54174.

Note: Supplementary data for this article are available at Cancer Epidemiology, Biomarkers and Prevention Online (http://cebp.aacrjournals.org/).

Requests for reprints: Robin J. Leach, Department of Cellular and Structural Biology, The University of Texas Health Science Center, 7703 Floyd Curl Drive, San Antonio, TX 78229-3900. Phone: 210-567-6947; Fax: 210-567-6781. E-mail: leach@uthscsa.edu

Copyright © 2009 American Association for Cancer Research.

doi:10.1158/1055-9965.EPI-09-0076

Downloaded from http://aacrjournals.org/cebp/article-pdf/18/6/1869/2246065/1869.pdf by guest on 15 August 2024

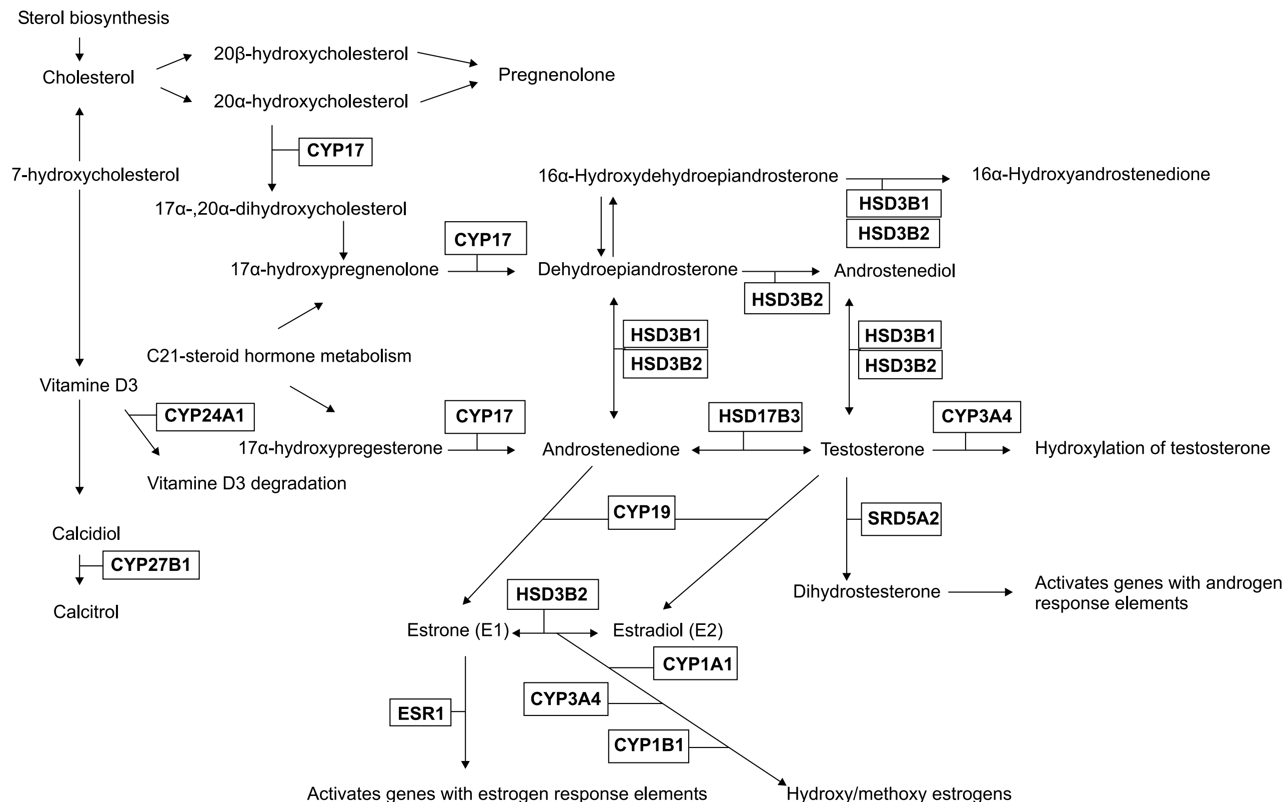

Figure 1. Summary of the steroid hormone pathway indicating the candidate genes studied in this report and the products from their enzymatic reactions.

(7q21.1, MIM 124010) into hydroxy and methoxy estrogens (Fig. 1).

The growth and differentiation of normal prostatic tissue is also promoted by interactions between the vitamin D and dihydrotestosterone pathways (9). Levels of the bioactive form of vitamin D are controlled by both the activating enzyme 1-A-hydroxylase (*CYP27B1*, 12q13.1-q13.3, MIM 609506) and the deactivating enzyme 24-hyroxylase (*CYP24A1*, 20q13, MIM 126065), which regulate cell growth and may reduce the risk of malignant transformation.

Functional variants in the genes of these pathways are likely to influence prostate carcinogenesis. Significant associations between variants in one or more of the above genes and risk for prostate cancer have been reported but with inconsistent findings. For example, meta-analyses of single nucleotide polymorphisms (SNPs) within *CYP17*, including rs743572, or SNPs rs523349 (V89L) and rs9282858 (A49T) within *SRD5A2* did not support evidence for association with risk of prostate cancer although smaller studies reported on significant risk effects (10-12). Recent efforts evaluating these risk variants relied on single-gene approaches. With a multicausal etiology of prostate cancer, it is likely that multiple risk variants act simultaneously, synergistically, or additively to influence prostate cancer susceptibility. Additionally, in most studies only a small proportion of the estimated number of genetic variants was analyzed, and the contributions of variants in regulatory, noncoding regions of genes, rather than in exons, were often omitted and probably underestimated because transcriptional regulatory elements are located in introns.

We systematically evaluated the association of genetic variants of 12 key genes within the steroid hormone pathway and prostate cancer risk, including *CYP17*, *HSD17B3*, *ESR1*, *SRD5A2*, *HSD3B1*, *HSD3B2*, *CYP19*, *CYP1A1*, *CYP1B1*, *CYP3A4*, *CYP27B1*, and *CYP24A1*, and genotyped 116 tagged SNPs covering the genes in 2,452 samples (886 cases and 1,566 controls) of non-Hispanic Caucasian, Hispanic Caucasian, or African American origin. We report on the association of individual SNPs as well as higher-order SNP-SNP interactions. We also tested the hypothesis that individuals carrying multiple copies of risk variants are at increased risk for prostate cancer.

Materials and Methods

Subjects. Study subjects included men from the San Antonio Center for Biomarkers of Risk of Prostate Cancer cohort. The San Antonio Center for Biomarkers of Risk of Prostate Cancer is funded by the National Cancer Institute and has been prospectively enrolling healthy male volunteers since 2001. On each annual visit, a digital rectal examination was done and the serum prostate-specific antigen level was determined. From this cohort, 231 incident cases (135 non-Hispanic Caucasians, 60 Hispanic Caucasians, and 36 African Americans) were available. We also included 655 cases with a known

Downloaded from http://aacrjournals.org/cebp/article-pdf/18/6/1869/2246065/1869.pdf by guest on 15 August 2024

history of prostate cancer that are enrolled within the same time period in a parallel study of prevalent prostate cancer. These prevalent cases were recruited through the same means and from the same metropolitan population as the San Antonio Center for Biomarkers of Risk of Prostate Cancer study and had a median time period of 3 y (range, 0-25 y) between disease diagnosis and enrollment into the study. Cases had biopsy-confirmed prostate cancer, and controls consisted of male volunteers at least 45 y old who had normal digital rectal examinations and prostate-specific antigen levels <2.5 ng/mL on all study visits. Race/ethnicity was self-reported on a questionnaire filled out at a clinic site at the time of recruiting, and the Hazuda algorithm (13) was used to classify race/ethnicity. All participants have established U.S. residency and are believed to be long-term residents. Study age among controls was the age at last follow-up, and age among cases was the age at prostate cancer diagnosis. Our control group was younger than our prostate cancer cases, with a mean age (SD) of 60.8 (8.8) y for the controls and a mean age (SD) of 65.5 (8.5) y for the cases ($P < 0.0001$). The clinical characteristics of the subjects are summarized in Table 1. A total of 1,452 non-Hispanic Caucasians (609 cases, 843 controls), 709 Hispanic Caucasians (195 cases, 514 controls), and 291 African Americans (82 cases, 209 controls) were included in this analysis. The San Antonio Center for Biomarkers of Risk of Prostate Cancer received Institutional Review Board approval, and informed consent was obtained from all subjects.

SNP Selection and Genotyping. DNA was isolated from participants' whole blood cells with the use of a QIAamp DNA Blood Maxi Kit (Qiagen).

A total of 120 SNPs within the 12 candidate genes were selected based on the following criteria. We first selected SNPs from available databases, the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/projects/SNP/) and SNPper (http://snpper.chip.org/), using the following criteria: (*a*) Within each gene, SNPs with a minor allele frequency >0.05, which leads to an amino acid substitution, and/or are in other coding regions of the gene and thus potentially functionally important were selected, (*b*) SNPs for which an association with prostate cancer has previously been shown as reported in literature were chosen. After this initial selection, we identified tagging SNPs within each gene using Haploview with the following criteria: (*a*) a minor allele frequency >0.05 to gain more statistical power, (*b*) an r^2 threshold of 0.8 and a log of odds threshold for multimarker testing of 3.0, (*c*) a minimum distance between tags of 60 bp, (*d*) we included our preselected SNPs (see above), (*e*) for each gene the search for SNPs extended to a 10 kb region surrounding the gene, and (*f*) we used the 2-marker and 3-marker haplotype tagging option (http://www.broad.mit.edu/mpg/haploview/). The selection was based on information on the European population as provided by HapMap (www.hapmap.org). The SNPs are described in Table 2.

Genotyping of 104 of the 120 SNPs was done with the GoldenGate assay of the VeraCode technology with the use of the BeadXpress Reader System according to the manufacturer's protocol (Illumina). SNPs rs2066479 (*HSD17B3*), rs1048943 (*CYP1A1*), rs2740574 (*CYP3A4*), and rs700519 (CYP19) were done with a standard TaqMan allelic discrimination assay with the use of an ABI 7900HT Sequence Detection System with the SDS 2.1 software (Applied Biosystems). Genotyping of the seven SNPs within *CYP1B1* was as previously described (14). For genotyping of rs2234693 and rs9340799 within *ESR1* and rs523349 and rs9282858 within *SRD5A2*, refer to Hernandez et al. (2006; ref. 15) and Torkko et al. (2008; ref. 16), respectively. SNP rs17115144 (*CYP17*) was analyzed by restriction fragment digestion of the amplified product as outlined by Lunn et al. (1999; ref. 17). Primers and probe sequences are available upon request. To ensure the reliability of the results, duplicate samples and/or known genotyped samples were included in the analysis as quality controls.

Statistics. Haploview version 4 beta 15 was used to check for Hardy-Weinberg equilibrium for each SNP and to measure linkage disequilibrium between the

Table 1. Clinical data of the study group

Subgroup	Cases (N = 886)	Controls (N = 1,566)	
	N (%)	N (%)	
Ethnic background			
Non-Hispanic Caucasian	609 (68.7%)	843 (53.8%)	
Hispanic Caucasian	195 (22.0%)	514 (32.8%)	
African American	82 (9.3%)	209 (13.4%)	
Age (y)			
≤50	36 (4.1%)	215 (13.7%)	
51-60	208 (23.6%)	581 (37.1%)	
61-70	389 (44.2%)	499 (31.9%)	
>70	248 (28.1%)	271 (17.3%)	
Mean ± SD	65.5 ± 8.5	60.8 ± 8.8	$P < 0.0001$
PSA (ng/mL)			
≤4.0	179	1,566	
4.1-10.0	38	0	
10.1-20.0	2	0	
>20.0	4	0	
Mean ± SD	3.1 ± 4.7	0.85 ± 0.44	$P < 0.0001$

Abbreviation: PSA, prostate-specific antigen.

Downloaded from http://aacrjournals.org/cebp/article-pdf/18/6/1869/2246065/1869.pdf by guest on 15 August 2024

Table 2. Genes, SNP selection, their location, and minor allele frequencies in cases and controls of each ethnicity

Gene	SNP	Position	SNP location/ AA change	Minor allele	Non-Hispanic Caucasians			Hispanic Caucasians			African Americans		
					MAF Cases	MAF Controls	P*	MAF Cases	MAF Controls	p*	MAF Cases	MAF Controls	p*
HSD3B2	rs1819698	chr1:119767042	3′UTR	A	0.142	0.104	**0.002**	0.173	0.14	0.140	0.518	0.525	0.894
	rs10923823	chr1:119789927		A	0.38	0.408	0.143 †	0.407	0.429	0.486 †	0.159	0.154	0.903
	rs1538989	chr1:119791376		A	0.487	0.43	**0.003**	0.428	0.378	0.099	0.817	0.836	0.591
	rs4659176	chr1:119791924		A	0.339	0.323	0.375	0.247	0.229	0.484	0.207	0.259	0.206
HSD3B1	rs947130	chr1:119818255		A	0.298	0.268	0.082	0.183	0.19	0.771	0.378	0.352	0.569
	rs6201	chr1:119855738	Val79Ile	A	0.499	0.5	0.964 †	0.5	0.5	1.000 †	0.433	0.427	0.910 †
	rs6428830	chr1:119856298		A	0.304	0.28	0.162	0.155	0.179	0.290	0.091	0.096	0.880
	rs6203	chr1:119858681	Leu338Leu	A	0.429	0.442	0.485	0.518	0.522	0.907	0.104	0.108	0.883
	rs2362819	chr1:119888607		C	0.327	0.306	0.239	0.246	0.262	0.556	0.549	0.554	0.916
SRD5A2	rs9332975	chr2:31603921	3′UTR	G	0.108	0.119	0.351	0.085	0.078	0.659	0.189	0.207	0.644
	rs2268794	chr2:31632908		A	0.135	0.139	0.749	0.095	0.081	0.423	0.329	0.333	0.928
	rs2268796	chr2:31635784		A	0.416	0.422	0.767	0.479	0.445	0.270	0.646	0.608	0.410
	rs2208532	chr2:31642493		G	0.439	0.427	0.496	0.503	0.52	0.566	0.402	0.333	0.132
	rs4952222	chr2:31653367		A	0.056	0.058	0.794	0.15	0.137	0.547	0.019	0.034	0.338
	rs523349	chr2:31659210	Val89Leu	C	0.271	0.284	0.553	0.388	0.346	0.274	0.343	0.296	0.355
	rs9282858	chr2:31659330	Ala49Thr	A	0.034	0.038	0.665	0.023	0.025	0.884	0	0.009	0.260
	rs9332960	chr2:31659458	Ter6Gln	A	0	0	0.000	0	0	0.000	0	0	0.000
	rs632148	chr2:31659535	5′UTR	X	failed	failed	failed	failed	failed	failed	failed	failed	failed
	rs3754838	chr2:31661804		G	0.113	0.116	0.825	0.083	0.074	0.560	0.112	0.099	0.641
CYP1B1	rs1800440	chr2:38151643	Ser453Asn	G	0.194	0.198	0.828	0.164	0.141	0.410	0.032	0.046	0.512
	rs1056836	chr2:38151707	Leu432Val	G	0.431	0.44	0.691	0.222	0.278	0.113	0.765	0.741	0.610
	rs1056827	chr2:38155681	Ser119Ala	A	0.277	0.278	0.949	0.316	0.321	0.913	0.52	0.44	0.239
	rs10012	chr2:38155894	Gly48Arg	G	0.283	0.273	0.630	0.288	0.301	0.719	0.515	0.458	0.303
	rs2617266	chr2:38156048		T	0.278	0.275	0.905	0.271	0.269	0.939	0.269	0.214	0.247
	rs2551188	chr2:38156298		A	0.281	0.272	0.683	0.288	0.298	0.783	0.493	0.431	0.258
	rs2567206	chr2:38157035		A	0.278	0.273	0.832	0.271	0.27	0.971	0.09	0.069	0.493
ESR1	rs2234693	chr6:152205028		G	0.476	0.487	0.639	0.379	0.399	0.629	0.534	0.491	0.447
	rs9340799	chr6:152205074		C	0.349	0.37	0.348	0.318	0.34	0.580	0.328	0.25	0.132
CYP3A4	rs4646450	chr7:99104254		A	0.16	0.154	0.640	0.309	0.314	0.863	0.854	0.852	0.958
	rs4646437	chr7:99203019		A	0.099	0.094	0.606	0.214	0.191	0.351	0.732	0.747	0.717
	rs2740574	chr7:99220032		C	0.028	0.031	0.711	0.074	0.091	0.585	0.629	0.665	0.602
HSD17B3	rs10989735	chr9:98017875		G	0.464	0.485	0.260	0.448	0.483	0.258	0.323	0.34	0.718
	rs10820020	chr9:98018481		G	0.215	0.226	0.476	0.24	0.232	0.757	0.079	0.105	0.364
	rs2479827	chr9:98028250		G	0.367	0.331	0.051	0.227	0.204	0.370	0.506	0.531	0.605
	rs2066479	chr9:98037631	Gly289Ser	A	0.063	0.039	**0.049**	0.049	0.044	0.835	0.081	0.125	0.342
	rs1810711	chr9:98048379		A	0.513	0.495	0.359	0.42	0.387	0.279	0.598	0.568	0.531
	rs2026001	chr9:98055916		A	0.35	0.371	0.255	0.487	0.466	0.488	0.262	0.231	0.454
	rs8190536	chr9:98056741		G	0.215	0.198	0.280	0.144	0.122	0.287	0.402	0.38	0.625
	rs2257157	chr9:98058263		G	0.459	0.432	0.158	0.34	0.388	0.108	0.439	0.488	0.309
	rs7039978	chr9:98061403		A	0.472	0.498	0.170	0.626	0.635	0.775	0.171	0.204	0.383
	rs7037932	chr9:98070492		A	0.351	0.327	0.216	0.214	0.206	0.768	0.375	0.444	0.179
	rs13302476	chr9:98075007		A	0.13	0.162	**0.018**	0.229	0.238	0.746	0.262	0.284	0.612
	rs280663	chr9:98089522		X	failed	failed	failed	failed	failed	failed	failed	failed	failed
	rs2479828	chr9:98090615		A	0.104	0.1	0.726	0.148	0.133	0.483	0.055	0.043	0.566
	rs9409407	chr9:98095070		A	0.123	0.101	0.065	0.047	0.055	0.572	0.019	0.015	0.801
	rs8190495	chr9:98101705		G	0.366	0.377	0.566	0.232	0.26	0.304	0.244	0.17	0.051
	rs2066474	chr9:98104246	5′UTR	G	0.168	0.187	0.197	0.224	0.275	0.061	0.188	0.18	0.844
	rs2026002	chr9:98124002		A	0.119	0.107	0.332	0.222	0.227	0.839	0.012	0.068	**0.007**
CYP17	rs284851	chr10:104568521		C	0.124	0.113	0.373	0.173	0.172	0.970	0.278	0.318	0.365
	rs619824	chr10:104571278		A	0.415	0.431	0.380	0.42	0.416	0.883	0.701	0.657	0.330
	rs10883782	chr10:104573922		G	0.145	0.178	**0.029**	0.109	0.112	0.880	0.083	0.116	0.281
	rs1004467	chr10:104584497		G	0.096	0.085	0.372	0.194	0.197	0.920	0.287	0.194	**0.043**
	rs6162	chr10:104586971	His46His	A	0.392	0.414	0.239	0.392	0.394	0.930	0.439	0.367	0.125
	rs743572	chr10:104587142	5′UTR	G	0.366	0.395	0.112	0.391	0.396	0.883	0.432	0.367	0.167
	rs17115144	chr10:104587298		C	0.345	0.361	0.567	0.381	0.412	0.578	0.5	0.331	**0.020**
	rs2486758	chr10:104587470		G	0.222	0.199	0.139	0.289	0.295	0.818	0.043	0.052	0.657
CYP27B1	rs1048691	chr12:56439215		A	0.21	0.218	0.641	0.33	0.327	0.931	0.329	0.42	0.053
	rs4646537	chr12:56443548		C	0.027	0.03	0.651	0.026	0.034	0.429	0.079	0.093	0.624
	rs4646536	chr12:56444255		G	0.31	0.326	0.359	0.343	0.323	0.501	0.262	0.293	0.472
	rs8176345	chr12:56444825	Leu314Leu	A	0.035	0.032	0.663	0.031	0.01	**0.010**	0	0.012	0.153
	rs10877012	chr12:56448352		X	failed	failed	failed	failed	failed	failed	failed	failed	failed
	rs703842	chr12:56449006	3′UTR	G	0.309	0.328	0.285	0.34	0.329	0.690	0.3	0.33	0.502
CYP19	rs9972359	chr15:49279146		G	0.417	0.414	0.871	0.402	0.383	0.528	0.768	0.759	0.825
	rs16964189	chr15:49281529		A	0.202	0.21	0.622	0.16	0.162	0.938	0.207	0.235	0.496
	rs934632	chr15:49283122		A	0.173	0.178	0.736	0.286	0.242	0.102	0.195	0.207	0.762
	rs934634	chr15:49287830		A	0.203	0.209	0.719	0.17	0.166	0.868	0.323	0.327	0.929
	rs2255192	chr15:49288127		A	0.203	0.209	0.709	0.173	0.164	0.712	0.323	0.336	0.769

(*Continued on the following page*)

Downloaded from http://aacrjournals.org/cebp/article-pdf/18/6/1869/2246065/1869.pdf by guest on 15 August 2024

Table 2. Genes, SNP selection, their location, and minor allele frequencies in cases and controls of each ethnicity (Cont'd)

Gene	SNP	Position	SNP location/ AA change	Minor allele	Non-Hispanic Caucasians			Hispanic Caucasians			African Americans		
					MAF Cases	MAF Controls	P*	MAF Cases	MAF Controls	p*	MAF Cases	MAF Controls	p*
	rs4646	chr15:49290136	3′UTR	A	0.24	0.254	0.384	0.448	0.424	0.420	0.256	0.275	0.662
	rs10046	chr15:49290278	3′UTR	A	0.557	0.536	0.270	0.374	0.408	0.257	0.317	0.265	0.232
	rs17601241	chr15:49295166		A	0.075	0.082	0.527	0.16	0.181	0.375	0.055	0.062	0.762
	rs700519	chr15:49295260	Arg264Cys	T	0.036	0.04	0.784	0.022	0.03	0.676	0.155	0.168	0.812
	rs2899472	chr15:49303347		A	0.281	0.26	0.213	0.124	0.188	**0.005**	0.055	0.04	0.458
	rs12439137	chr15:49303596		G	0.153	0.112	**0.002**	0.194	0.217	0.381	0.058	0.085	0.310
	rs12592697	chr15:49312465		A	0.33	0.347	0.345	0.402	0.376	0.388	0.305	0.318	0.770
	rs700518	chr15:49316404	Val80Val	G	0.544	0.517	0.160	0.343	0.385	0.154	0.28	0.216	0.114
	rs10519295	chr15:49319939		G	0.111	0.108	0.784	0.026	0.055	**0.025**	0.006	0.009	0.714
	rs10459592	chr15:49323433		A	0.418	0.428	0.579	0.523	0.463	0.053	0.61	0.605	0.918
	rs11636686	chr15:49329358		A	0.534	0.51	0.201	0.325	0.37	0.126	0.159	0.123	0.285
	rs12050772	chr15:49332163		C	0.493	0.468	0.178	0.281	0.328	0.100	0.183	0.151	0.369
	rs2414099	chr15:49336074		G	0.138	0.151	0.327	0.258	0.258	0.984	0.165	0.225	0.117
	rs11636639	chr15:49350384		C	0.481	0.453	0.130	0.303	0.366	**0.032**	0.53	0.41	**0.012**
	rs2470176	chr15:49371231		G	0.139	0.146	0.607	0.186	0.156	0.208	0.524	0.525	0.995
	rs2470158	chr15:49375687		A	0.093	0.098	0.607	0.075	0.074	0.953	0.116	0.12	0.884
	rs17523880	chr15:49379835		A	0.147	0.134	0.314	0.088	0.106	0.334	0.043	0.025	0.277
	rs2470152	chr15:49382264		G	0.551	0.524	0.167	0.407	0.43	0.457	0.451	0.457	0.907
	rs17523922	chr15:49386497		G	0.097	0.102	0.664	0.064	0.075	0.506	0.018	0.037	0.257
	rs3751592	chr15:49393870		G	0.303	0.266	**0.042** †	0.369	0.372	0.942	0.291	0.224	0.143
	rs3751591	chr15:49394002		G	0.166	0.16	0.659	0.169	0.189	0.414	0.043	0.049	0.742
	rs1902584	chr15:49398946		A	0.082	0.08	0.908	0.131	0.122	0.650	0.037	0.077	0.083
	rs2445762	chr15:49405000		G	0.264	0.247	0.315	0.397	0.379	0.556	0.427	0.401	0.587
	rs7168331	chr15:49408275		G	0.384	0.362	0.241	0.49	0.485	0.873	0.579	0.525	0.253
	rs7174997	chr15:49409420		A	0.168	0.192	0.092	0.157	0.164	0.763	0.037	0.062	0.243
	rs868475	chr15:49423746		G	0.05	0.059	0.281	0.12	0.112	0.693	0.299	0.256	0.317
	rs2470164	chr15:49425853		C	0.566	0.501	**0.003**	0.392	0.389	0.911	0.131	0.056	**0.027**
	rs2445773	chr15:49430233		A	0.174	0.181	0.633	0.276	0.261	0.585	0.207	0.213	0.885
CYP1A1	rs2198843	chr15:72788283		G	0.166	0.147	0.233	0.389	0.44	0.676	0.5	0.546	0.853
	rs1048943	chr15:72800038	Val462Ile	G	0.045	0.037	0.490	0.279	0.278	0.993	0.033	0.011	0.239
	rs4646421	chr15:72803245		A	0.162	0.238	0.185	0.643	0.473	**0.013**	0.7	0.493	0.083
	rs2470893	chr15:72806502		A	0.304	0.316	0.504	0.127	0.121	0.778	0.049	0.053	0.850
CYP24A1	rs2762929	chr20:52199602		G	0.389	0.399	0.601	0.369	0.387	0.531	0.683	0.725	0.329
	rs8118441	chr20:52199792		T	0.139	0.142	0.837	0.186	0.186	0.994	0.311	0.278	0.445
	rs6068810	chr20:52202758		A	0.055	0.045	0.251	0.034	0.019	0.128	0.043	0.025	0.277
	rs6097807	chr20:52202862		G	0.243	0.255	0.472	0.268	0.247	0.432	0.488	0.448	0.399
	rs4809957	chr20:52204578	3′UTR	G	0.215	0.212	0.816	0.247	0.232	0.547	0.438	0.388	0.283
	rs2762934	chr20:52204668	3′UTR	A	0.192	0.187	0.779	0.134	0.162	0.210	0.146	0.179	0.362
	rs1570669	chr20:52207834		G	0.323	0.337	0.421	0.312	0.3	0.684	0.604	0.58	0.620
	rs2296239	chr20:52208935	Pro375Pro	A	0.204	0.213	0.557	0.247	0.233	0.580	0.439	0.392	0.318
	rs6068816	chr20:52214498	Thr248Thr	A	0.108	0.109	0.948	0.106	0.079	0.130	0.067	0.046	0.334
	rs4809958	chr20:52215845		C	0.156	0.152	0.794	0.16	0.12	0.056	0.085	0.068	0.486
	rs3787554	chr20:52216087		A	0.093	0.091	0.795	0.085	0.045	**0.005**	0.03	0.037	0.709
	rs2244719	chr20:52216265		G	0.451	0.484	0.088	0.552	0.569	0.596	0.244	0.262	0.659
	rs2762941	chr20:52217059		A	0.388	0.376	0.505	0.296	0.319	0.434	0.506	0.531	0.605
	rs2181874	chr20:52217885		A	0.256	0.241	0.355	0.253	0.289	0.192	0.36	0.423	0.179
	rs4809960	chr20:52219480		G	0.232	0.237	0.738	0.242	0.209	0.198	0.11	0.154	0.179
	rs2296241	chr20:52219626	Ala184Ala	G	0.484	0.47	0.447	0.554	0.559	0.887	0.549	0.497	0.279
	rs2245153	chr20:52219813		G	0.181	0.194	0.390	0.18	0.131	**0.027**	0.315	0.289	0.563
	rs2585428	chr20:52220304		A	0.449	0.473	0.194	0.351	0.352	0.949	0.433	0.503	0.143
	rs13038432	chr20:52220709		G	0.084	0.083	0.919	0.052	0.045	0.606	0.012	0.022	0.466
	rs6022999	chr20:52221420		G	0.241	0.239	0.930	0.225	0.233	0.782	0.594	0.627	0.485
	rs2248359	chr20:52224925		A	0.396	0.396	0.991	0.36	0.343	0.566	0.646	0.627	0.668

NOTE: Significant P values are in bold.
Abbreviation: MAF, minor allele frequency.
*Assumes Hardy-Weinberg equilibrium.
†Markers not in Hardy-Weinberg equilibrium ($P < 0.01$) in both the controls and the cases.

SNPs in the controls and cases of each race/ethnicity (ref. 18; http://www.broad.mit.edu/mpg/haploview/).

The allele frequency for each SNP was determined in each ethnic group, and the frequencies among the case-control groups were compared with the use of the χ^2 test. Association analyses were stratified by ethnicity and done with the R statistical software version 2.5.1. The odds ratio (OR) and its 95% confidence interval (CI) were estimated by unconditional logistic regression as a measure of the associations between genotypes and prostate cancer risk. We tested for additive, dominant, and recessive associations. To correct for multiple testing, we used the method of Storey and Tibshirani (2003) based on the concept of false discovery rate (19). This estimation showed that, for $P < 0.01$, the probability that the association is expected to be a true positive in our

Downloaded from http://aacrjournals.org/cebp/article-pdf/18/6/1869/2246065/1869.pdf by guest on 15 August 2024

sample group is >50%. To estimate the independent effect of a significant SNP while adjusting for other SNPs, we used a generalized linear model function from the R statistical package so that all SNPs are entered into a single multivariate logistic regression model. SNPs in this model were taken to have additive effects.

The random forest (RF) algorithm, implemented in R (http://www.r-project.org/) was applied to assess the

Table 3. Significant results from individual SNP effects on prostate cancer in non-Hispanic Caucasians, Hispanic Caucasians, and African Americans after correction for multiple testing

Gene	SNP	Non-Hispanic Caucasians					P
		Genotype	Controls (N)	Cases (N)	OR*	95% CI	
HSD3B2	rs1819698	GG	673	439	1.00	—	
		AA	9	15	2.66	1.13-6.24	0.025
		AG	157	138	1.34	1.03-1.74	0.029
		AA/AG vs GG	166	153	1.41	1.09-1.82	0.008
		AA vs AG/GG	9	15	2.50	1.07-5.85	0.035
		A# vs G#	839	592	1.41	1.12-1.77	0.003
HSD3B2	rs1538989	GG	272	157	1.00	—	
		AA	155	142	1.55	1.15-2.11	0.005
		AG	412	294	1.20	0.93-1.54	0.160
		AA/AG vs GG	567	436	1.29	1.02-1.64	0.033
		AA vs AG/GG	155	142	1.39	1.07-1.80	0.014
		A# vs G#	839	593	1.24	1.07-1.45	0.005
CYP19	rs12439137	AA	645	362	1.00	—	
		GG	9	6	1.10	0.38-3.20	0.863
		AG	165	145	1.56	1.20-2.03	0.001
		GG/AG vs AA	174	151	1.53	1.19-1.99	0.001
		GG vs AG/AA	9	6	0.99	0.34-2.87	0.979
		G# vs A#	819	513	1.45	1.14-1.84	0.003
CYP19	rs2470152	GG	238	166	1.00	—	
		AA	197	106	0.75	0.55-1.02	0.067
		AG	404	322	1.16	0.90-1.49	0.257
		AA/AG vs GG	601	428	1.02	0.80-1.29	0.873
		AA vs AG/GG	197	106	0.68	0.52-0.89	0.005
		A# vs G#	839	594	0.87	0.76-1.03	0.121
CYP19	rs2470164 †	CC	127	191	1.00	—	
		AA	126	114	0.58	0.41-0.82	0.002
		AC	206	282	0.91	0.68-1.21	0.512
		AA/AC vs CC	332	396	0.78	0.60-1.03	0.076
		AA vs AC/CC	126	114	0.62	0.46-0.83	0.001
		A# vs C#	459	587	0.77	0.65-0.91	0.003
Gene	**SNP**	**Hispanic Caucasians**					**P**
		Genotype	**Controls (N)**	**Cases (N)**	**OR***	**95% CI**	
CYP19	rs934632	GG	228	93	1.00	—	
		AA	25	10	1.03	0.46-2.31	0.939
		AG	140	91	1.71	1.17-2.51	0.006
		AA/AG vs GG	165	101	1.61	1.11-2.32	0.011
		AA vs AG/GG	25	10	0.82	0.37-1.79	0.613
		A# vs G#	393	194	1.33	0.99-1.78	0.061
CYP19	rs10519295	AA	350	185	1.00	—	
		GG	0	1	ND	ND	0.978
		AG	43	8	0.34	0.15-0.76	0.009
		GG/AG vs AA	43	9	0.39	0.18-0.84	0.016
		GG vs AG/AA	0	1	ND	ND	0.978
		G# vs A#	393	194	0.45	0.22-0.94	0.034
CYP19	rs10459592	CC	120	38	1.00	—	
		AA	91	47	1.82	1.06-3.12	0.029
		AC	182	109	2.03	1.28-3.22	0.003
		AA/AC vs CC	273	156	1.96	1.27-3.05	0.003
		AA vs AC/CC	91	47	1.13	0.73-1.73	0.585
		A# vs C#	393	194	1.34	1.04-1.74	0.026
CYP24A1	rs3787554 †	GG	358	162	1.00	—	
		AA	0	1	ND	ND	0.979
		AG	35	31	2.05	1.18-3.56	0.011
		AA/AG vs GG	35	32	2.12	1.23-3.66	0.007
		AA vs AG/GG	0	1	ND	ND	0.979
		A# vs G#	393	194	2.14	1.25-3.66	0.005

(*Continued on the following page*)

Downloaded from http://aacrjournals.org/cebp/article-pdf/18/6/1869/2246065/1869.pdf by guest on 15 August 2024

Table 3. Significant results from individual SNP effects on prostate cancer in non-Hispanic Caucasians, Hispanic Caucasians, and African Americans after correction for multiple testing (Cont'd)

Gene	SNP	African Americans					P
		Genotype	Controls (*N*)	Cases (*N*)	OR*	95% CI	
CYP19	rs3751592	AA	72	40	1.00	—	
		GG	13	7	0.83	0.29-2.35	0.725
		AG	22	32	2.55	1.29-5.06	0.008
		GG/AG vs AA	35	39	1.90	1.03-3.51	0.040
		GG vs AG/AA	13	7	0.61	0.22-1.67	0.330
		G# vs A#	107	79	1.26	0.81-1.96	0.298

Abbreviation: ND, not determined.
*Age-adjusted ORs from unconditional logistic regression analyses.
†SNPs with main effect independent from other significant SNPs.

importance of the polymorphisms and potential nonlinear interactions with respect to the outcome variable (20). For each RF analysis, the number of trees to grow (''ntree'') was set to 5,000, and the number of predictors to be randomly selected at each tree node (''mtry'') was set to 20. To better characterize and validate the ''important'' polymorphisms identified by both the RF algorithm as well as single SNP analysis, and to determine possible interactions between the top ranked or significant SNPs within the same or different genes, we used the generalized multifactor dimensionality reduction (GMDR) method (ref. 21; The GMDR *P* values and other significance tests are representative of consistency with RF but do not indicate independent confirmation of RF findings). An exhaustive search of all possible *n* models was done for subsets of selected SNPs (with *n* between 1 and 10). We used 10-fold cross-validations for each set of *n* SNPs, and the model with the lowest misclassification error was selected. *P* values were determined by the sign test, a robust nonparametric test implemented in the software. We also calculated empirical *P* values for interaction based on 1,000 permutations. Missing values were imputed for both the RF and GMDR methods. The RF method imputes missing data by computing a weighted average of subjects with completed data, in which the weights are similarity measures to the subjects with missing data. The MDR data tool was used to impute missing values in the GMDR analysis (http://sourceforge.net/project/showfiles.php?group_id=131001&package_id=144548).

The cumulative effect of combined genotypes on prostate cancer risk was estimated by counting the number of genotypes associated with prostate cancer on the basis of the best-fitting genetic inheritance from single SNP analysis. ORs and their 95% CIs were calculated for men carrying any combination of one, two, or more alleles associated with prostate cancer compared with men carrying none of the risk genotypes with the use of unconditional logistic regression analysis. We selected SNPs that were not in linkage disequilibrium with each other (D' < 0.8). If several SNPs presented higher linkage disequilibrium values, we chose a SNP in a coding region above an intronic SNP and also selected the most significant SNP.

For all statistical analyses, age (modeled as a continuous predictor) was used as covariate. Individuals with missing data for a particular analysis were removed from the analysis. All statistical tests were two-sided, and significance was set at $P < 0.05$.

A hypothetical function for SNPs that were significant for single SNP analysis and/or selected by RF was assessed with the use of in silico analysis of transcription factor binding sites: both possible alleles of each SNP were tested for their binding capability to human transcription factors with the use of the web tool 'transcription element search system' (TESS; ref. 22; http://www.cbil.upenn.edu/cgi-bin/tess/tess). The options used were 21 bases of genomic sequence around each SNP (10 bases on either side of the SNP) and a string-based search query with default settings. A log-likelihood score of >16 for a pretty good match (deficit ≤1.0) and >18 for a mismatch (deficit >1.0) were used as cut-off values for reporting.

Results

Single SNP Analysis. A total of 120 SNPs were genotyped in 2,452 samples. The majority of SNPs (>80%) were successfully genotyped in >90% of the samples. Three SNPs failed, and one SNP was not polymorphic in our sample. Table 2 gives the minor allele frequencies of the SNPs estimated in each of the three ethnicities in our study sample. Differences in allele frequency distribution among the ethnicities were found for several SNPs (e.g., rs4646450 in *CYP3A4* shows a minor allele frequency of ~15% for the A allele in non-Hispanic Caucasians, of ~30% in Hispanic Caucasians, and of ~85% in African Americans). Significant case/control differences of allele frequencies at a level <0.05 were observed for eight polymorphisms in non-Hispanic Caucasians, for seven SNPs within Hispanic Caucasians, and for five SNPs in African Americans (Table 2). One SNP, rs6201, showed deviation from Hardy-Weinberg equilibrium ($P < 0.01$) in the cases and controls of all three ethnic study groups. In non-Hispanic and Hispanic Caucasians, rs10923823 was also not in Hardy-Weinberg equilibrium in the cases nor the controls, and variant rs3751592 was out of Hardy-Weinberg equilibrium in non-Hispanic Caucasians. Although the error rate was <0.2%, SNPs that were not in Hardy-Weinberg equilibrium were left out for further statistical analyses in the respective study groups.

Logistic regression analysis was done for all SNPs in the context of recessive, dominant, and additive models after adjustment for age. Thirteen SNPs in four genes were significantly associated with prostate cancer risk in non-Hispanic Caucasians at the $P < 0.05$ level. For Hispanic Caucasians, 19 SNPs in six genes were found to be significant for prostate cancer risk, and in African

Downloaded from http://aacrjournals.org/cebp/article-pdf/18/6/1869/2246065/1869.pdf by guest on 15 August 2024

Table 4. Cumulative effects of risk variants

Markers	No. of risk genotypes	Controls	Cases	OR (95% CI)*	P
Non-Hispanic Caucasians					
rs1819698, rs12439137, rs2470152 †	0	174	76	Ref	—
	1	394	245	1.39 (1.0-1.9)	0.044
	2	221	154	1.56 (1.11-2.20)	0.011
	3	30	38	2.87 (1.64-5.02)	0.0002
	Trend			2.20 (1.44-3.38)	0.0003
Hispanic Caucasians					
rs10459592, s3787554 ‡	0	107	32	Ref	—
	1	264	136	1.88 (1.17-3.02)	0.009
	2	22	26	4.58 (2.19-9.61)	6.1×10^{-5}
	Trend			4.29 (2.11-8.72)	0.00006

*Age-adjusted.
†rs2470164 was not included because it is in complete linkage disequilibrium with rs2470152.
‡rs934632 and rs10519295 were not included because they are in complete linkage disequilibrium with rs10459592.

Americans, 5 SNPs in four genes showed significant results. The estimated ORs and 95% CIs for significant results in all three ethnicities are shown in Supplementary Tables S1, S2, and S3. After correction for multiple testing, estimating a false discovery rate of <50% for $P < 0.01$, five SNPs remained significant in non-Hispanic Caucasians, rs1819698 and rs1538989 within *HSD3B2*, rs12439137, and rs2470152 and rs2470164 within *CYP19* (P = 0.001-0.008; Table 3). There was a moderate increase (~1.5) or decrease (~0.6) in the ORs associated with these significant results. In Hispanic Caucasians, four SNPs remained significant after correction for multiple testing (P = 0.003-0.009), three SNPs within *CYP19* (rs934632, rs10519295, and rs10459592) and one SNP within *CYP24A11* (rs3787554). The P values were associated with a >2-fold increase or decrease in OR (Table 3). In African Americans, the association with prostate cancer risk remained significant for rs3751592 in *CYP19*, with P = 0.008 and a >2.5 increase in OR for the AG heterozygote carriers. In all three sample groups, one or multiple single SNPs within *CYP19* showed significant association with prostate cancer. Except for rs1819698 (*CYP19*), located within the 3′ untranslated region of the gene, all other significant SNPs are within intronic regions of the genes.

We also tested whether the statistically significant associations with prostate cancer represented independent risk factors in our sample groups. In the non-Hispanic Caucasians, rs2470164 remained significant after correction for the other four significant SNPs (P = 0.0007) under the additive model. SNP rs3787554 in the Hispanic Caucasians remained significant after adjusting for the other significant SNPs (P = 0.004; data not shown). In African Americans, only one SNP remained significant after correction for multiple testing.

Cumulative Effect of Significant SNPs. To evaluate possible cumulative effects of the risk alleles defined in the single SNP analysis, we did an age-adjusted multivariate logistic regression on combinations of at-risk alleles compared with the no-risk allele reference. In non-Hispanic Caucasians, the combination of the three risk alleles not in linkage disequilibrium with each other showed a significant association with prostate cancer ($P_{\text{trend}} = 2.9 \times 10^{-4}$), and a >2-fold increase in risk (OR, 2.20; 95% CI, 1.44-3.38) was observed for individuals carrying all three risk alleles compared with individuals without any risk allele (Table 4). A similar observation was found in the Hispanic Caucasians, in which the significant association between the combination of the risk alleles of two SNPs and prostate cancer increases the risk significantly (OR, 4.29; 95% CI, 2.11-8.72; $P_{\text{trend}} = 6.4 \times 10^{-5}$). In African Americans, there was only one SNP that remained significant after correction for multiple testing.

RF Analysis. To classify and predict the underlying relationships between the predictor variables and prostate cancer, we assessed the importance of the polymorphisms studied in this report by running the RF algorithm. For non-Hispanic Caucasians, Hispanic Caucasians, and African Americans, the number of SNPs chosen were 92, 91, and 97, respectively, corresponding to the missing genotypes in each sample group. Supplementary Table S4 (*b*) shows the top-12 ranked SNPs as calculated by RF in each of the three sample groups (a choice of 12 SNPs is arbitrary albeit motivated by reasoning that we investigated 12 genes in this report). Of the 5 significant SNPs from single SNP analysis in non-Hispanic Caucasians, 2 are ranked within the first 12 SNPs by RF. In Hispanic Caucasians, 2 of the significant SNPs were selected in the 12 most important SNPs. In African Americans, only 1 of the 3 significant SNPs was scored by RF within the 12 most important SNPs. One SNP, rs11636639 (*CYP19*), was common in the selection by RF in all three race/ethnic groups. SNP rs2470152 (*CYP19*) was present in the top-12 list of both non-Hispanic and Hispanic Caucasians. Two SNPs, rs10519295 (*CYP19*) and rs8176345 (*CYP27B1*), were within the top 12 SNPs for Hispanic Caucasians and African Americans. Although several of the SNPs selected by RF are coding SNPs, no nonsynonymous SNPs were selected.

SNP-SNP Interaction. The RF algorithm is aimed at detecting the variables that have an important role in the prediction of the outcome, but it does not indicate how these predictors act (i.e., through main effects or through interactive effects involving other variables). Furthermore, because the genes studied are all found in one pathway, possible interactions between single SNPs might exist. SNP-SNP interaction was determined with the use of GMDR. For each racial/ethnicity group, we

Downloaded from http://aacrjournals.org/cebp/article-pdf/18/6/1869/2246065/1869.pdf by guest on 15 August 2024

Table 5. Best multigenic models, testing accuracies, cross-validation consistencies, and *P* values identified by GMDR

SNPs into the model	Ethnicity	Best model	Testing accuracy	CVC	*P**	*P*†	Genes involved
13 significant‡	Non-Hispanic Caucasians	rs1538989-rs2479827-rs17523880-rs2470164	0.63	10/10	0.001	<0.001	*HSD3B2, HSD17B3, CYP19*
19 significant†	Hispanic Caucasians	rs523349-rs1810711-rs2066474-rs4646-rs10459592-rs2470152-rs2762941	0.60	10/10	0.001	0.005	*SRD5A2, HSB17B3, CYP19, CYP24A1*
5 significant†	African American	rs10012-rs17115144-rs1048691-rs11636639-rs3751592	0.60	10/10	0.05	0.04	*CYP1B1, CYP17, CYP27B1, CYP19*
Top 12 RF	Hispanic Caucasians	rs10989735-rs1810711-rs10459592-rs11636639-rs2470152-rs868475-rs2245153	0.61	10/10	0.01	0.001	*HSD17B3, CYP19, CYP24A1*
Top 12 RF	African American	rs2208532-rs2026002-rs6162-rs8176345-rs12439137-rs2414099-rs11636639-rs1902584-rs7174997-rs13038432	0.63	10/10	0.05	0.014	*SRD5A2, HSD17B3, CYP17, CYP27B1, CYP19, CYP24A1*

**P* value from the sign test in GMDR.
†*P* value from the permutation test.
‡Significant before multiple correction.

looked at possible interactions between the significant SNPs found in the single SNP analysis before multiple correction as well as the top-12 ranked SNPs from RF. A summary of the significant results for the models that have minimum prediction error (maximum testing accuracy) and/or maximum cross-validation consistency (CVC) is presented in Table 5.

Considering SNPs that are significant for single SNP analysis before multiple correction (see Supplementary Tables S1, S2, and S3), we found a significant interaction between SNPs rs1538989, rs2479827, rs17523880, and rs2470164 within *HSD3B2, HSD17B3*, and *CYP19* in the non-Hispanic Caucasians (permutation $P < 0.001$; CVC, 10/10; testing accuracy, 0.63), suggesting that the four SNPs not only have a main effect, some of which are marginal, but also show a joint action of the three genes. In Hispanic Caucasians, a seven-locus model for rs523349, rs1810711, rs2066474, rs4646, rs10459592, rs2470152, and rs2762941, involving genes *SRD5A2, HSB17B3, CYP19*, and *CYP24A1*, showed a significant interaction (permutation $P = 0.005$; CVC, 10/10; testing accuracy, 0.60). A marginal significant interaction between five SNPs, rs10012, rs17115144, rs1048691, rs11636639, and rs3751592, within the genes *CYP1B1, CYP17, CYP27B1*, and *CYP19* is found in African Americans (permutation $P = 0.04$; CVC, 10/10; testing accuracy, 0.60), indicating that those SNPs act jointly to confer prostate cancer risk.

Considering the 12 top-ranked SNPs by RF in non-Hispanic Caucasians, no significant interactions were found (data not shown). In Hispanic Caucasians, a highly significant interaction (permutation $P = 0.001$) was found for 7 of the top-12 ranked SNPs with a CVC of 10/10 and a testing accuracy of 0.61. In African Americans, the model with the highest CVC and maximum testing accuracy included 10 of the 12 top-ranked SNPs (permutation $P = 0.014$; CVC, 8/10; testing accuracy, 0.63).

Alterations of Binding Sites for Transcription Factors. To investigate possible functional implications of the significant SNPs as well as the SNPs identified by RF, a search for transcription factor binding sites with the use of the transcription element search system was done. The motive behind the analysis is that recent findings show that transcription factor binding sites may be commonly located in introns or other noncoding regions of the genome (23). Three of the five SNPs, rs1538989, rs2470152, and rs2470164, that show significant association with prostate cancer in non-Hispanic Caucasians; one significant SNP, rs3787554, in Hispanics; and the significant SNP rs3751592 in African Americans have changes in transcription factor binding properties related to allelic alterations (Supplementary Table S4, *a*). The presence of the C allele in rs1538989 (*HSD3B2*) creates a binding site for a transcriptional activator, PU.1. The Maz binding site, which is a transcriptional activator, is present for the T allele in rs2470152 (*CYP19*). A binding site for nuclear factor κB, activated by the tumor necrosis factor gene and also a transcriptional repressor of the H-2Kb gene in metastatic tumor cells, is present for the G allele in rs2470164 (*CYP19*). SNP rs3787554 (*CYP24A1*) has an Sp1 binding site for the G allele. The A allele of rs3751592 (*CYP19*) has a binding site for a transcriptional modulator of tumor necrosis factor-α.

Downloaded from http://aacrjournals.org/cebp/article-pdf/18/6/1869/2246065/1869.pdf by guest on 15 August 2024

Considering the SNPs that were prioritized by RF, 4 of the 12 SNPs (rs4809957, rs1538989, rs2470152, and rs703842) were found to display a binding difference of transcription factors in an allele-specific manner in non-Hispanic Caucasians; 4 of the 12 SNPs (rs11636686, rs10989735, rs1810711, and rs2470152) affect a binding site for transcription factors in Hispanic Caucasians; and 2 SNPs (rs2414099 and rs2208532) show allele-specific alterations of binding motives in African Americans (Supplementary Table S4, *b*). Of particular interest is the finding for rs4809957 (*CYP24A1*) in non-Hispanic Caucasians, for which the A allele creates a binding site for the retinoic acid–responsive element. In Hispanic Caucasians, 1 SNP is noteworthy: rs10989735 (*HSD17B3*), in which the G allele contains an Sp1 transcription factor binding site. In African Americans, the G allele at rs2208532 (*SRD5A2*) has a LEF1 binding site, which induces neoplastic transformation.

Discussion

Prostate carcinogenesis is a multistep process involving a multifactorial interplay between genetic and environmental factors; as a result, the effect of individual SNPs are unlikely to be substantial and may be of limited value in predicting risk. A comprehensive approach is required, such as the pathway-based multigenic method integrating multiple polymorphisms that interact in the same pathway. Combining multiple low-penetrant to modestly penetrant SNPs may amplify predictive power.

To estimate prostate cancer risk conferred by individual SNPs, SNP-SNP interactions, and/or cumulative SNP effects, we selected 12 key genes of the steroid hormone pathway and report on the contribution of 116 SNPs within those genes to prostate cancer susceptibility. A total of 886 cases and 1,566 healthy controls from three racial/ethnic groups of the San Antonio Center for Biomarkers of Risk of Prostate Cancer cohort were genotyped.

When analyzed individually, a limited number of SNPs conferred increased risk of prostate cancer after correction for multiple testing (five in non-Hispanic Caucasians, four in Hispanic Caucasians, and three in African Americans). SNPs within *CYP19* were significantly associated with prostate cancer in all three ethnic groups, indicating that this gene likely plays a consistent role in prostate cancer susceptibility. *CYP19* is a key enzyme in the conversion of androstenedione and testosterone into estrone and estradiol, and a lack of normal function could affect testosterone concentrations, in turn influencing risk of prostate cancer (24, 25). Furthermore, rs2470164 (*CYP19*) is significant in both non-Hispanic Caucasians and African Americans. Importantly, the allele frequency distribution is significantly different between the two ethnicities with C as the major allele in non-Hispanic Caucasians, whereas this allele is the rare allele in African Americans, and the effect of the SNP differs between both races; male homozygotes of the AA allele have a decreased risk in non-Hispanic Caucasians, whereas heterozygote AC increases risk in African Americans. The SNP rs2470164 has not been reported in previous association studies with prostate cancer.

Of the SNPs analyzed in this report, 35 (~30%) have been previously studied in prostate cancer. Only two of the significant SNPs from single SNP analysis (rs1819698 in *HSD3B2* and rs2470152 in *CYP19*, both significant in non-Hispanic Caucasians) have been previously reported and showed no association with prostate cancer in non-Hispanic Whites (26, 27). Our findings for SNPs within *SRD5A2* are consistent with a previous meta-analysis that did not find evidence for a significant main effect of rs523349 (V89L) or rs9282858 (A49T; ref. 12). However, our results suggest the involvement of *SRD5A2* in a polygenic model for prostate cancer risk; in particular, rs523349 and rs2208532 are part of a multilocus model that shows significant interaction in Hispanic Caucasians and African Americans, respectively. We also confirmed previous findings from a meta-analysis by Ntais et al. (10) and Setawian et al. (11) that do not support a significant role of variants within *CYP17* in prostate cancer susceptibility in non-Hispanic Caucasians. However, consistent with the findings of Ntais et al. (10), our results of SNP-SNP interaction also suggest that variants within *CYP17* may play a role in susceptibility to prostate cancer among African Americans. We could not confirm previous significant findings for association with prostate cancer of coding SNPs found in Caucasians, Hispanics, or African Americans, in particular the nonsynonymous SNPs within *SRD5A2* (rs523349 and rs9282858; refs. 2, 28-32), *CYP1B1* (rs1800440, rs1056836, and rs1056827; refs. 33-35), *HSD17B3* (rs2066479; ref. 36), and *CYP1A1* (rs1048943; ref. 37). Explanations for these discrepancies include between-study variability in sample size, diagnostic criteria used, and/or ethnic background of the sample studied.

Higher-order SNP-SNP and/or gene-gene interactions among the reported genetic variants indicate that the combined effect of the polymorphisms with or without significant main effects confer risk for prostate cancer. For each race/ethnicity, significant interactions were found between SNPs with main effects (some marginal) for single SNP analysis. These results suggest interactions among *HSD3B2, HSD17B3*, and *CYP19* in non-Hispanic Caucasians; among *SRD5A2, HSB17B3, CYP19*, and *CYP24A1* in Hispanic Caucasians; and among *HSD3B2, CYP1B1, CYP17, CYP27B1*, and *CYP19* in African Americans. Interactions between SNPs selected by RF, several without a main effect, confirm that complex polygenic models are related to prostate cancer risk. These findings show the importance of SNPs that may not have significance at the individual level and which may be overlooked without an interaction analysis. It should be mentioned that the testing accuracies calculated by GMDR, between 60% and 63%, represent an improvement over the rate of 50% expected under random prediction and shows the importance of multigenic models.

Caution should be used when interpreting these results, especially about the biological significance of these multilocus epistatic models. The statistical modeling of interactions may also not correspond to a true biological interaction (38, 39). However, based on the transcription element search system web tool, hypotheses related to molecular mechanisms resulting in differential activity of genes can be speculated. Several SNPs found to be important in this study, either from single SNP analysis or selected by RF method, have an allele-specific change in transcriptional binding sites and could contribute to transcriptional regulation of oncogenes, as in the case of rs2208532 (*SRD5A2*), or tumor necrosis factors, as in the case of rs2470164 (*CYP19*) or

Downloaded from http://aacrjournals.org/cebp/article-pdf/18/6/1869/2246065/1869.pdf by guest on 15 August 2024

rs3751592 (*CYP19*). Further studies may identify the effect of these allele-specific changes and their contribution to prostate carcinogenesis.

Our results indicate a significant synergistic effect for increasing numbers of potential high-risk genotypes in both non-Hispanic and Hispanic Caucasians. Similar findings for a cumulative effect, in which commonly occurring SNPs incrementally contribute to prostate cancer risk, have been shown for genes from the steroid hormone pathway (30, 40).

The GMDR method has many advantages over others, such as the ability to adjust for covariates and applicability to population-based study designs, including unbalanced case-control samples. In addition, GMDR does not rely on binary splits but does a systematic search through all possible genotype combinations of the SNP variables and may reveal more interactions than methods based on, for example, classification and regression trees.

A limitation of the study is that race/ethnicity is self-reported and, as in any association study, population stratification might lead to confounding results. However, misclassifications are equally likely in cases and controls, and should not have a substantial impact on the outcome. In addition, the majority of markers were not significant, suggesting that population stratification does not represent a significant problem. Another limitation is that we only evaluated the trait prostate cancer using the covariate age. Other endophenotypes and/or interactions with nongenetic factors, as well as the risk factor family history may be essential and could be related to the risk of prostate cancer, which may be missed in this study. For this report, we selected 12 key genes from the steroid hormone pathway. Because the exact mechanism by which these genes might act in prostate cancer development and progression is largely unknown, the genes investigated may represent only a fraction of all relevant genes. Moreover, the SNPs selected in the study are not an exhaustive list of all existing common SNPs and may not fully represent the genetic variability of these genes. However, a careful selection of genes and SNPs was made based on previous functional analysis and/or a previous association study related to prostate cancer risk and, thus, high-priority genes and SNPs were included in the analysis.

In summary, our results support the hypothesis that prostate cancer develops through a variety of interconnected genes, some of which belong to the steroid hormone pathway, and that there are synergistic interactions among these SNPs related to increasing prostate cancer risk. This finding is consistent with the polygenic model for cancer susceptibility, in which the combined effect of variants in many genes, each conferring a small or modest increase in risk, together account for a substantial fraction of the risk for prostate cancer. Our approach of evaluating a large number of genetic variants in a large sample size is one of the first efforts in exploring the effect of high-order gene-gene interactions on prostate cancer risk and might help in improving the ability and accuracy in assessing individual risk of prostate cancer.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

We thank all the study subjects in the San Antonio Center for Biomarkers of Risk of Prostate Cancer and in the prevalent prostate cancer studies at The University of Texas Health Science Center at San Antonio for their participation, and the San Antonio Center for Biomarkers of Risk of Prostate Cancer clinical staff for their skilled assistance. We utilized the Institutional Genomic Resource Core for genotyping with the Illumina system.

References

1. Schaid DJ. The complex genetic epidemiology of prostate cancer. Hum Mol Genet 2004;13 Spec No 1:R103–21.
2. Makridakis NM, Reichardt JK. Molecular epidemiology of hormone-metabolic loci in prostate cancer. Epidemiol Rev 2001;23:24–9.
3. Hsing AW. Hormones and prostate cancer: what's next? Epidemiol Rev 2001;23:42–58.
4. Cheng E, Lee C, Grayhack J. Endocrinology of the prostate. In: Lepor H, Lawson RK, editors. Prostate diseases. Philadelphia (PA): Saunders; 1993. p. 57–71.
5. Waterman MR, Keeney DS. Genes involved in androgen biosynthesis and the male phenotype. Horm Res 1992;38:217–21.
6. Van Den Akker EL, Koper JW, Boehmer AL, et al. Differential inhibition of 17α-hydroxylase and 17,20-lyase activities by three novel missense CYP17 mutations identified in patients with P450c17 deficiency. J Clin Endocrinol Metab 2002;87:5714–21.
7. Kagimoto M, Winter JS, Kagimoto K, Simpson ER, Waterman MR. Structural characterization of normal and mutant human steroid 17 α-hydroxylase genes: molecular basis of one example of combined 17 α-hydroxylase/17,20 lyase deficiency. Mol Endocrinol 1988;2:564–70.
8. Chung BC, Picado-Leonard J, Haniu M, et al. Cytochrome P450c17 (steroid 17 α-hydroxylase/17,20 lyase): cloning of human adrenal and testis cDNAs indicates the same gene is expressed in both tissues. Proc Natl Acad Sci U S A 1987;84:407–11.
9. Leman ES, DeMiguel F, Gao AC, Getzenberg RH. Regulation of androgen and vitamin d receptors by 1,25-dihydroxyvitamin D3 in human prostate epithelial and stromal cells. J Urol 2003;170:235–40.
10. Ntais C, Polycarpou A, Ioannidis JP. Association of the CYP17 gene polymorphism with the risk of prostate cancer: a meta-analysis. Cancer Epidemiol Biomarkers Prev 2003;12:120–6.
11. Setiawan VW, Schumacher FR, Haiman CA, et al. CYP17 genetic variation and risk of breast and prostate cancer from the National Cancer Institute Breast and Prostate Cancer Cohort Consortium (BPC3). Cancer Epidemiol Biomarkers Prev 2007;16:2237–46.
12. Ntais C, Polycarpou A, Ioannidis JP. SRD5A2 gene polymorphisms and the risk of prostate cancer: a meta-analysis. Cancer Epidemiol Biomarkers Prev 2003;12:618–24.
13. Hazuda HP, Comeaux PJ, Stern MP, Haffner SM, Eifler CW, Rosenthal M. A comparison of three indicators for identifying Mexican Americans in epidemiologic research. Methodological findings from the San Antonio Heart Study. Am J Epidemiol 1986;123:96–112.
14. Beuten J, Gelfond JA, Byrne JJ, et al. CYP1B1 variants are associated with prostate cancer in non-Hispanic and Hispanic Caucasians. Carcinogenesis 2008;29:1751–7.
15. Hernandez J, Balic I, Johnson-Pais TL, et al. Association between an estrogen receptor α gene polymorphism and the risk of prostate cancer in black men. J Urol 2006;175:523–7.
16. Torkko KC, van Bokhoven A, Mai P, et al. VDR and SRD5A2 polymorphisms combine to increase risk for prostate cancer in both non-Hispanic White and Hispanic White men. Clin Cancer Res 2008;14:3223–9.
17. Lunn RM, Bell DA, Mohler JL, Taylor JA. Prostate cancer risk and polymorphism in 17 hydroxylase (CYP17) and steroid reductase (SRD5A2). Carcinogenesis 1999;20:1727–31.
18. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 2005;21:263–5.
19. Storey JD, Tibshirani R. Statistical significance for genomewide studies. Proc Natl Acad Sci U S A 2003;100:9440–5.
20. Breiman L. Random forests. Machine learning 2001;45:5–32.
21. Lou XY, Chen GB, Yan L, et al. A generalized combinatorial approach for detecting gene-by-gene and gene-by-environment

Downloaded from http://aacrjournals.org/cebp/article-pdf/18/6/1869/2246065/1869.pdf by guest on 15 August 2024

interactions with application to nicotine dependence. Am J Hum Genet 2007;80:1125–37.
22. Schug JaOGC. TESS: Transcription element search software on the WWW. Technical report CBIL-TR-1997-1001-v0.0 Computational Biology and Informatics Laboratory School of Medicine. PA: University of Pennsylvania; 1997.
23. Long F, Liu H, Hahn C, Sumazin P, Zhang MQ, Zilberstein A. Genome-wide prediction and analysis of function-specific transcription factor binding sites. In Silico Biol 2004;4:395–410.
24. Singh PB, Matanhelia SS, Martin FL. A potential paradox in prostate adenocarcinoma progression: oestrogen as the initiating driver. Eur J Cancer 2008;44:928–36.
25. Miner MM, Seftel AD. Testosterone and ageing: what have we learned since the Institute of Medicine report and what lies ahead? Int J Clin Pract 2007;61:622–32.
26. Douglas JA, Zuhlke KA, Beebe-Dimmer J, et al. Identifying susceptibility genes for prostate cancer-a family-based association study of polymorphisms in CYP17, CYP19, CYP11A1, and LH-β. Cancer Epidemiol Biomarkers Prev 2005;14:2035–9.
27. Cunningham JM, Hebbring SJ, McDonnell SK, et al. Evaluation of genetic variations in the androgen and estrogen metabolic pathways as risk factors for sporadic and familial prostate cancer. Cancer Epidemiol Biomarkers Prev 2007;16:969–78.
28. Forrest MS, Edwards SM, Houlston R, et al. Association between hormonal genetic polymorphisms and early-onset prostate cancer. Prostate Cancer Prostatic Dis 2005;8:95–102.
29. Cicek MS, Conti DV, Curran A, et al. Association of prostate cancer risk and aggressiveness to androgen pathway genes: SRD5A2, CYP17, and the AR. Prostate 2004;59:69–76.
30. Cussenot O, Azzouzi AR, Nicolaiew N, et al. Combination of polymorphisms from genes related to estrogen metabolism and risk of prostate cancers: the hidden face of estrogens. J Clin Oncol 2007;25: 3596–602.
31. Salam MT, Ursin G, Skinner EC, Dessissa T, Reichardt JK. Associations between polymorphisms in the steroid 5-α reductase type II (SRD5A2) gene and benign prostatic hyperplasia and prostate cancer. Urol Oncol 2005;23:246–53.
32. Nam RK, Toi A, Vesprini D, et al. V89L polymorphism of type-2, 5-α reductase enzyme gene predicts prostate cancer presence and progression. Urology 2001;57:199–204.
33. Tang YM, Green BL, Chen GF, et al. Human CYP1B1 Leu432Val gene polymorphism: ethnic distribution in African-Americans, Caucasians and Chinese; oestradiol hydroxylase activity; and distribution in prostate cancer cases and controls. Pharmacogenetics 2000;10:761–6.
34. Cicek MS, Liu X, Casey G, Witte JS. Role of androgen metabolism genes CYP1B1, PSA/KLK3, and CYP11α in prostate cancer risk and aggressiveness. Cancer Epidemiol Biomarkers Prev 2005;14:2173–7.
35. Chang BL, Zheng SL, Isaacs SD, et al. Polymorphisms in the CYP1B1 gene are associated with increased risk of prostate cancer. Br J Cancer 2003;89:1524–9.
36. Margiotti K, Kim E, Pearce CL, Spera E, Novelli G, Reichardt JK. Association of the G289S single nucleotide polymorphism in the HSD17B3 gene with prostate cancer in Italian men. Prostate 2002;53: 65–8.
37. Chang BL, Zheng SL, Isaacs SD, et al. Polymorphisms in the CYP1A1 gene are associated with prostate cancer risk. Int J Cancer 2003;106: 375–8.
38. Siemiatycki J, Thomas DC. Biological models and statistical interactions: an example from multistage carcinogenesis. Int J Epidemiol 1981;10:383–7.
39. Greenland S. Modeling and variable selection in epidemiologic analysis. Am J Public Health 1989;79:340–9.
40. Sobti RC, Gupta L, Singh SK, Seth A, Kaur P, Thakur H. Role of hormonal genes and risk of prostate cancer: gene-gene interactions in a North Indian population. Cancer Genet Cytogenet 2008;185:78–85.
41. Pongubala JM, Nagulapalli S, Klemsz MJ, McKercher SR, Maki RA, Atchison ML. PU.1 recruits a second nuclear factor to a site important for immunoglobulin κ3′ enhancer activity. Mol Cell Biol 1992;12:368–78.
42. Pongubala JM, Van Beveren C, Nagulapalli S, et al. Effect of PU.1 phosphorylation on interaction with NF-EM5 and transcriptional activation. Science 1993;259:1622–5.
43. Kennedy GC, Rutter WJ. Pur-1, a zinc-finger protein that binds to purine-rich sequences, transactivates an insulin promoter in heterologous cells. Proc Natl Acad Sci U S A 1992;89:11498–502.
44. Plaksin D, Baeuerle PA, Eisenbach L. KBF1 (p50 NF-κB homodimer) acts as a repressor of H-2Kb gene expression in metastatic tumor cells. J Exp Med 1993;177:1651–62.
45. Hohmann HP, Brockhaus M, Baeuerle PA, Remy R, Kolbeck R, van Loon AP. Expression of the types A and B tumor necrosis factor (TNF) receptors is independently regulated, and both receptors mediate activation of the transcription factor NF-κB. TNFα is not needed for induction of a biological effect via TNF receptors. J Biol Chem 1990;265:22409–17.
46. Ishii S, Kadonaga JT, Tjian R, Brady JN, Merlino GT, Pastan I. Binding of the Sp1 transcription factor by the human Harvey ras1 proto-oncogene promoter. Science 1986;232:1410–3.
47. Shannon MF, Gamble JR, Vadas MA. Nuclear proteins interacting with the promoter region of the human granulocyte/macrophage colony-stimulating factor gene. Proc Natl Acad Sci U S A 1988;85:674–8.
48. Kliewer SA, Umesono K, Mangelsdorf DJ, Evans RM. Retinoid X receptor interacts with nuclear receptors in retinoic acid, thyroid hormone and vitamin D3 signalling. Nature 1992;355:446–9.
49. Qian X, Samadani U, Porcella A, Costa RH. Decreased expression of hepatocyte nuclear factor 3 α during the acute-phase response influences transthyretin gene transcription. Mol Cell Biol 1995;15: 1364–76.
50. Peers B, Voz ML, Monget P, et al. Regulatory elements controlling pituitary-specific expression of the human prolactin gene. Mol Cell Biol 1990;10:4690–700.
51. Cai RL. Human CART1, a paired-class homeodomain protein, activates transcription through palindromic binding sites. Biochem Biophys Res Commun 1998;250:305–11.
52. Anderson GM, Freytag SO. Synergistic activation of a human promoter *in vivo* by transcription factor Sp1. Mol Cell Biol 1991;11: 1935–43.
53. Pascal E, Tjian R. Different activation domains of Sp1 govern formation of multimers and mediate transcriptional synergism. Genes Dev 1991;5:1646–56.
54. Lee JS, Galvin KM, Shi Y. Evidence for physical interaction between the zinc-finger transcription factors YY1 and Sp1. Proc Natl Acad Sci U S A 1993;90:6145–9.
55. Kataoka H, Joh T, Miura Y, et al. AT motif binding factor 1-A (ATBF1-A) negatively regulates transcription of the aminopeptidase N gene in the crypt-villus axis of small intestine. Biochem Biophys Res Commun 2000;267:91–5.
56. Quong MW, Massari ME, Zwart R, Murre C. A new transcriptional-activation motif restricted to a class of helix-loop-helix proteins is functionally conserved in both yeast and mammalian cells. Mol Cell Biol 1993;13:792–800.
57. Aoki M, Hecht A, Kruse U, Kemler R, Vogt PK. Nuclear endpoint of Wnt signaling: neoplastic transformation induced by transactivating lymphoid-enhancing factor 1. Proc Natl Acad Sci U S A 1999; 96:139–44.

Downloaded from http://aacrjournals.org/cebp/article-pdf/18/6/1869/2246065/1869.pdf by guest on 15 August 2024