A Comprehensive Analysis of Common \textit{IGF1}, \textit{IGFBP1} and \textit{IGFBP3} Genetic Variation with Prospective IGF-I and IGFBP-3 Blood Levels and Prostate Cancer Risk Among Caucasians

Fredrick R. Schumacher (1), Iona Cheng (2), Matthew L. Freedman (3), Lorelei Mucci (4,7), Naomi E. Allen (10), Michael N. Pollak (11), Richard B. Hayes (12), Daniel O. Stram (1), Federico Canzian (13), Brian E. Henderson (1), David J. Hunter (6,7), Jarmo Virtamo (15), Jonas Manjer (16), J. Michael Gaziano (8,9), Laurence N. Kolonel (2), Anne Tjønneland (17), Demetris Albanes (18), Eugenia E. Calle (19), Edward Giovannucci (4,5,7), E. David Crawford (20), Christopher A. Haiman (1), Peter Kraft (6), Walter C. Willett (4,5), Michael J. Thun (19), Loïc Le Marchand (2), Rudolf Kaaks (14), Heather Spencer Feigelson (19,21), H. Bas Bueno-de-Mesquita (22), Domenico Palli (23), Elio Riboli (24), Eiliv Lund (25), Pilar Amiano (22), Gerald Andriole (26), Alison M. Dunning (27), Dimitrios Trichopoulos (28), Meir J. Stampfer (4,7), Timothy J. Key (10), Jing Ma(7),*

*Corresponding Author: Jing Ma, Channing Laboratory, Brigham and Women’s Hospital, 181 Longwood Avenue, Boston, MA 02115. Phone: 617-525-2708; Fax: 617-525-2008; E-mail: jing.ma@channing.harvard.edu

Author Affiliations:
(1) Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA 90033, USA
(2) Epidemiology Program, Cancer Research Center, University of Hawaii, Honolulu, HI 96813, USA
(3) Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA 02115, USA
(4) Department of Epidemiology, Harvard School of Public Health, Boston, MA 02115, USA
(5) Department of Nutrition, Harvard School of Public Health, Boston, MA 02115, USA
(6) Program in Molecular and Genetic Epidemiology, Harvard School of Public Health, Boston, MA 02115, USA
(7) Channing Laboratory, Brigham and Women’s Hospital, Harvard Medical School, Boston, MA 02115, USA
(8) Division of Aging, Brigham and Women’s Hospital, Harvard Medical School, Boston, MA 02115, USA
(9) MAVERIC, VA Boston Healthcare System, Boston, MA 02115, USA
(10) Cancer Epidemiology Unit, University of Oxford, Oxford OX3 7BN, UK
(11) Cancer Prevention Research Unit, Department of Medicine and Oncology, Lady Davis Research Institute of Jewish General Hospital and McGill University, Montreal, Quebec H3T 1E2, Canada
(12) New York University School of Medicine, NY, NY 10016, USA
(13) Genomic Epidemiology Group, German Cancer Research Centre, Heidelberg 69120, Germany
(14) Division of Cancer Epidemiology, German Cancer Research Centre, Heidelberg 69120, Germany
(15) Department of Health Promotion and Chronic Disease Prevention, National Public Health Institute, Helsinki FIN-00300, Finland
(16) Department of Surgery, Malmö University Hospital, Malmö S-205 02, Sweden
(17) Institute of Cancer Epidemiology, Danish Cancer Society, Copenhagen DK-2100, Denmark
(18) Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, MD 20892, USA
(19) Department of Epidemiology and Surveillance Research, American Cancer Society, Atlanta, GA 30303, USA
(20) Urologic Oncology, University of Colorado at Denver Health Sciences Center, Denver, CO 80045, USA
(21) Kaiser Permanente, Denver 80237, CO, USA
(22) National Institute for Public Health and the Environment (RIVM), Bilthoven 3720 BA, The Netherlands
(23) Molecular and Nutritional Epidemiology Unit, ISPO-Cancer Research and Prevention Institute, Florence 50139, Italy
(24) Department of Epidemiology and Public Health, Imperial College, London SW7 2AZ, United Kingdom
(25) Insitute of Community Medicine, University of Tromsö, Tromsö 9019, Norway
(26) Divison of Urologic Surgery, Washington University School of Medicine, St. Louis, MO 63110, USA
(27) Department of Oncology, University of Cambridge, Cambridge CB2 1TN, United Kingdom
(28) Department of Hygiene and Epidemiology, School of Medicine, University of Athens, Athens 11527, Greece
Abstract

The insulin-like growth factor pathway (IGF) has been implicated in prostate development and carcinogenesis. We conducted a comprehensive analysis, utilizing a resequencing and tagging SNP approach, between common genetic variation in the $IGF1$, $IGF$ binding protein ($BP$) 1 and $IGFBP3$ genes with IGF-I and IGFBP-3 blood levels, and prostate cancer (PCa) risk, among Caucasians in the NCI Breast and Prostate Cancer Cohort Consortium. We genotyped fourteen $IGF1$ single-nucleotide polymorphisms (SNPs) and sixteen $IGFBP1$/IGFBP3 SNPs to capture common (MAF ≥5%) variation among Caucasians. For each SNP, we assessed the geometric mean difference in IGF blood levels (N=5,684) across genotypes and the association with PCa risk (6,012 PCa cases/6,641 controls). We present two-sided statistical tests and correct for multiple comparisons. A nonsynonymous $IGFBP3$ SNP in exon1, rs2854746 (Gly32Ala), was associated with IGFBP-3 blood levels ($P_{adj}=8.8x10^{-43}$) after adjusting for the previously established $IGFBP3$ promoter polymorphism A-202C (rs2854744); IGFBP-3 blood levels were 6.3% higher for each minor allele. For $IGF1$ SNP rs4764695, the risk estimates among heterozygotes was 1.01 (99%CI: 0.90-1.14) and 1.20 (99%CI: 1.06-1.37) for variant homozygotes with overall PCa risk. The corrected allelic P-value was $8.7x10^{-3}$. IGF-I levels were significantly associated with PCa risk ($P_{trend}=0.02$) with a 21% increase of PCa risk comparing the highest quartile to the lowest quartile. We have identified SNPs significantly associated with IGFBP-3 blood levels, but none of these alter PCa risk; however a novel $IGF1$ SNP, not associated with IGF-I blood levels, shows preliminary evidence for association with PCa risk among Caucasians.
Introduction

The role of the insulin-like growth factor (IGF) pathway has been studied extensively in both normal and transformed cells. Both in vivo (1-3) and in vitro (4-6) studies demonstrate that IGF-I binding to the IGF type 1 receptor modulates cellular proliferation, differentiation and apoptosis – important characteristics in tumorigenesis (7-11). Circulating levels of IGF-I derive mostly from the liver; over 90% is complexed with IGF-binding protein 3 (IGFBP-3) and an acid-labile subunit thus reducing bioavailability (12, 13). However, many types of tissues, including certain neoplasms (10), are capable of producing IGF-I locally. Although the main effect of IGFBP-3 is thought to be inhibition of cell growth and proliferation due to sequestration of the IGF-I ligand, recent research suggests IGFBP-3 has antiproliferative and proapoptotic effects independent of IGF-I (14, 15).

Elevated blood levels of IGF-I have been associated with several cancers, most commonly with prostate cancer (PCa), although later studies have found weaker associations than initially reported (16-23). A recent meta-analysis of twelve prospective studies reported a 38% increased risk of developing PCa when comparing the highest to lowest quartile of IGF-I levels (24). Although circulating IGFBP-3 levels were inversely associated with PCa risk in earlier studies, recent findings have been mostly null (23-27).

Nutrition remains a key determinant of circulating IGF-I levels (28, 29), but heritability studies have estimated that the proportion of variance explained by inherited genetic variation ranges from 38% to >80% for IGF-I and IGFBP-3 blood levels (30-34). The specific genetic variants that contribute to heritable risk are not well defined. Results between an upstream IGF1 repeat sequence (CA)n and IGF-I blood levels have varied (35-38), most studies, including a recently published meta-analysis (39), have reported a null association (40-44). In a case-control study, Johansson et al. recently reported a marginally significant association ($P=0.02$) between an IGF1 haplotype, previously reported associated with PCa risk (45), and IGF-I blood levels among controls in the Cancer Prostate in Sweden (CAPS) study (46); however, the authors were unable to replicate this haplotype-IGF-I association in a prospective study. In contrast, the significant decreased IGFBP-3 blood levels among carriers of the C allele of the IGFBP3 A-202C promoter polymorphism (rs2854744) has been observed across multiple studies of both men and women (39, 41, 43, 47-54). Additionally, an in vitro transient
transfection assay demonstrated the C allele had 50% lower activity than the A allele (55).

Most previous genetic studies of the IGF1 locus with PCa risk have focused on an upstream IGF1 (CA)n repeat with equivocal results (56-60). Recently, two investigations comprehensively examined IGF1 genetic variation with PCa risk selecting SNPs by public databases (i.e. HapMap), exonic resequencing, or both. In the Multietnic Cohort (MEC) study Cheng et al. identified an association with an upstream IGF1 SNP (rs7965399) and PCa (61), while in the CAPS study, Johannsson et al. used public databases reported a significant increased risk of PCa for single-copy carriers of an IGF1 haplotype spanning intron 2 through the 3’UTR (45). The relationship between common genetic variation in the IGFBPs and PCa risk has only been thoroughly examined by Cheng et al. in the MEC; no association was found between IGFBP1 and IGFBP3 polymorphisms with PCa risk (62).

We conducted a comprehensive haplotype tag-SNP analysis of the common genetic variation in IGF1, IGFBP1 and IGFBP3 in relation to IGF-I and IGFBP-3 blood levels and PCa risk among Caucasians in the NCI Breast and Prostate Cancer Cohort Consortium (BPC3), a pooled nested case-control study from seven cohorts (63). The large sample size of the BPC3 having 6,012 prospective PCa cases and 6,641 controls allows us to detect modest genetic effects and assess effect modification. In addition, we are able to examine risk in clinically important subgroups of PCa defined by stage and Gleason score at diagnosis.

Results

Characteristics of the studies within the BPC3 are presented in Table 1. Briefly, the cases and controls were comparable across cohorts with respect to demographic and other potentially PCa-related factors, with the exception of height where the ATBC Study (Finnish population) and EPIC cohorts (8 European countries) are shorter in stature. Family history was not available for the EPIC and PHS cohorts. PCa clinical information such as stage (63%) and Gleason score (56%) was available for over half of the cases.

The IGF1 locus was characterized by four haplotype blocks (Figure 1, bottom panel). The IGFBP1 and IGFBP3 loci are near each other, separated by 19-kb, and were characterized by three haplotype blocks defined by 12 htSNPs (Figure 2- bottom panel). We included four additional IGFBP3 SNPs. The average genotyping
success rate across cohorts was 94.7% (ranging from 84.4% to 99.8%). No deviation from Hardy-Weinberg equilibrium was observed for any SNP (at the $P<0.01$ level) among controls across more than one study. The controls minor allele frequencies were similar across the cohorts, albeit the ATBC (Finnish) cohort differed slightly for a few markers ($IGF1$: Supplementary Table 1; $IGFBP1$ & $IGFBP3$: Supplementary Table 2).

**$IGF1$ Genetic Variation and IGF-I Blood Levels**

Individual cohort and pooled values of log-transformed IGF-I blood levels were similar as were the case and control levels within each cohort (Supplementary Figure 1). In a pooled analysis, 3 of 14 htSNPs were nominally associated with IGF-I blood levels (Figure 1: green triangles): rs35767 (Block 1; $P_{uncorr}=7.9 \times 10^{-3}$), rs12821878 (Block 2; $P_{uncorr}=1.0 \times 10^{-3}$) and rs1549593 (Block 3; $P_{uncorr}=3.3 \times 10^{-3}$). As shown in Figure 1, the pair-wise LD for these three markers is negligible except between rs12821878 and rs1549593 ($r^2=0.27$ among PLCO controls). The geometric means and 95%CIs by SNP and haplotypes are presented in Supplementary Tables 3 and 4. After correcting for multiple-comparisons, none of the $IGF1$ markers associations remained statistically significant ($P_{corr}>0.07$) with IGF-I blood levels.

**$IGFBP1/IGFBP3$ Genetic Variation and IGFBP-3 Blood Levels**

The distributions of IGFBP-3 blood levels were not as uniform as IGF-I blood levels across the cohorts (Supplementary Figure 2). Within each cohort, case and control blood levels of IGFBP-3 were very similar. The geometric means and 95%CIs by SNP and haplotypes are presented in Supplementary Table 5 and 6.

Among the 12 htSNPs and 4 candidate SNPs in $IGFBP1$ and $IGFBP3$, six were significantly associated with blood levels before and after adjusting for multiple-comparisons (Figure 2: green triangles). For the most strongly associated htSNP, rs2854746 ($P_{corr}=8.8 \times 10^{-43}$), the mean IGFBP-3 blood level was 3,046 ng/ml for wild-type homozygotes, 3,263 ng/ml for heterozygotes and 3,442 ng/ml for variant homozygotes. The extensively studied $IGFBP3$ promoter polymorphism rs2854744 (A-202C) was also strongly associated with IGFBP-3 blood levels in the univariate analysis ($P_{corr}=8.1 \times 10^{-34}$). However, after simultaneously including all six significant $IGFBP3$ SNPs in a multi-SNP linear regression, only rs2854746 remained statistically
significant overall \((P_{uncorr}=1.4x10^{-10})\) and among controls \((P_{uncorr}=4.9x10^{-8})\), while rs2854744 (A-202C) was no longer associated with IGFBP-3 blood levels \((P=0.91)\). Stratified analysis of the IGFBP3 htSNPs rs2854746 and rs2854744 (A-202C) showed a consistent result such that across the three genotypes for rs2854746, the mean IGFBP-3 blood levels remained unchanged by rs2854744 (A-202C) genotypes (Table 2). However, the differences in mean IGFBP-3 blood levels were statistically significant across rs2854746 in the heterozygote (CA) and variant homozygote (AA) strata defined by rs2854744 (A-202C) genotypes \((P_{CA}=2.7x10^{-6}; P_{AA}=4.1x10^{-6})\).

**IGF1 Genetic Variation and Prostate Cancer Risk**

The associations between the fourteen IGF1 htSNPs and overall PCa risk among Caucasians are presented in the upper panel of Figure 1 (red circles). The tests for heterogeneity (Supplementary Table 7) across cohorts were not statistically significant, thus we present the pooled results for the main effect analyses. We found nominally significant associations between IGF1 SNPs rs2373722 \((P_{uncorr}=2.0x10^{-5})\) and rs4764695 \((P_{uncorr}=1.2x10^{-4})\) with overall PCa risk (Supplementary Table 8). After controlling for multiple comparisons, the corrected p-value for rs2373722 was no longer significant \((P_{corr}=0.14)\), whereas rs4764695 remained statistically significant \((P_{corr}=8.7x10^{-3})\). For rs4764695, the overall risk estimate was 1.01 (99%CI: 0.90-1.14) for heterozygotes and 1.20 (99%CI: 1.06-1.37) for variant homozygotes, with consistent point estimates for heterozygotes and variant homozygotes across the cohorts (Figure 3). The results remained unchanged when previously published data from the MEC were excluded from the analysis \((P_{uncorr}=2.8x10^{-5})\;\text{(Supplementary Table 8)}. IGF1 marker rs4764695 was not significantly associated with IGF-I blood levels. Testing for effect modification by several variables of interest (family history of PCa, age at diagnosis, BMI, and height) revealed no statistically significant heterogeneity in any of the subgroup analyses for the 14 htSNPs (data not shown).

We examined the risk association of each SNP stratified by stage at diagnosis (High Stage: C or D; Low Stage: A or B) and Gleason score at diagnosis (High Grade: ≥8; Low Grade: <8) compared to controls (data not shown). Among the fourteen IGF1 SNPs, rs4764695 was the only marker remaining statistically significant \((P_{uncorr}<0.01)\) across all of the strata except for stage (Table 3). The risk estimate was slightly
greater for high grade cancer (Gleason score ≥8, OR=1.43) than low grade cancer (Gleason score 2-7, OR=1.19) for variant homozygotes, but the test of heterogeneity was only marginally statistically significant ($P=0.065$).

The cohort-specific haplotype frequencies for $IGF1$ among Caucasians in the MEC panel and other six cohorts are shown in Supplementary Table 9 and the pooled and cohort-specific associations of these haplotypes and overall PCa risk are presented in Supplementary Table 10. The tests for heterogeneity by cohort across the haplotype blocks were not statistically significant. We observed statistically significant global $p$-values for $IGF1$ block 3 ($P_{uncorr}=0.002$) and block 4 ($P_{uncorr}=0.004$). Since SNP rs2373722 resides within block 3 and rs4764695 resides within block 4, we removed both SNPs from their respective blocks and found that neither block remained statistically significant with PCa risk (data not shown), suggesting that the haplotype analysis did not provide any additional insight beyond the SNP analysis. None of the haplotype analyses were significant for effect modification by family history, age at diagnosis, BMI, height, or PCa diagnostic variables (data not shown).

**IGFBP1 & IGFBP3 Genetic Variation and Prostate Cancer Risk**

Figure 2 shows the results for overall PCa risk for the twelve htSNPs and four additional $IGFBP1$ and $IGFBP3$ SNPs (red circles). None of the 16 SNPs were nominally associated with PCa risk at the $P_{uncorr}<0.01$ level. The $IGFBP3$ markers significantly associated with blood levels were not associated with PCa risk. The cohort-specific results are presented in Supplementary Table 11 and the $IGFBP1$ and $IGFBP3$ SNP analyses with and without the previously reported MEC samples are available in Supplementary Table 12. None of the tests for heterogeneity (by cohort) or effect modification (family history, age at diagnosis, BMI, and height in tertiles and quartiles) were statistically significant in any of the subgroup analyses (data not shown).

The haplotype frequencies within the MEC panel and by cohort are presented in Supplementary Table 13. Similar to the $IGFBP1$ and $IGFBP3$ SNP analyses, neither the global tests for haplotype blocks nor the test for any individual haplotype were statistically significant (Supplementary Table 14) in the pooled or cohort-specific analyses.
Discussion

Substantial epidemiologic and experimental evidence exist implicating the IGF pathway in prostate carcinogenesis. Although many studies have demonstrated an increased risk of PCA, especially advanced disease among men with high IGF-I blood levels (16-18, 27), comprehensive genetic analyses of the IGF pathway with PCA risk are limited. In this large consortium (n=6,012 PCA cases) from seven prospective studies, we conducted a thorough analysis of common genetic variation of three primary loci in the IGF pathway ($IGF1$, $IGFBP1$ and $IGFBP3$) and observed among Caucasians an association between an $IGF1$ SNP (rs4764695, MAF=0.49, ~34-kb downstream of $IGF1$ exon 4) and PCA risk following multiple testing corrections. Men carrying the homozygous variant, had a 20% higher risk of developing PCA (OR=1.20; 99%CI: 1.06-1.37; $P_{adj}=8.69\times10^{-3}$) compared to those with the wild-type controlling for age in five year intervals, cohort, and country of residence for EPIC.

The $IGF1$ htSNPs presented here are a subset of the $IGF1$ htSNPs reported in a study by the MEC. The MEC authors did not find a significant association between PCA risk and rs4764695 ($P>0.30$) but they had limited power among whites (23%) for this marker (61). They reported a marginally significant association between PCA and heterozygotes for an $IGF1$ htSNP (rs7965399) located in the 5’ region (OR$_{het}=1.25$; 95%CI: 1.09-1.43), whereas we observed a null association between rs7965399 and PCA with ($P=0.334$) and without ($P=0.689$) the MEC Caucasian participants. The majority of the MEC nested PCA case-control study, however, is non-Caucasian (African American=28%, Latino=28%, Japanese=20%, native Hawaiian=3%) and the different risk associations among overlapping $IGF1$ htSNPs suggest either racial/ethnic differences within this region, the causative SNP remains unknown, or false-positive findings. The $IGF1$ htSNP results for African Americans presented by Cheng et al. remained unchanged when we included an additional 105 African American PCA cases from PLCO (data not shown).

The Swedish CAPS case-control study (N=2,863 PCA cases, all Caucasian) also evaluated $IGF1$ genetic variation with PCA by genotyping SNPs identified in the HapMap Phase I data (45). A marginally significant association was reported for carriers of one copy of a haplotype spanning from intron 2 to the 3’ UTR of $IGF1$ with PCA risk ($P=0.02$ adjusting for multiple-testing) corresponding to a similar region covered
by our \textit{IGF1} haplotype block 3. Although the three SNPs in the CAPS haplotype (rs2033178, rs7136446, and rs6220) were not included in our \textit{IGF1} htSNP panel, our \textit{IGF1} htSNP rs2373722 is in perfect LD ($r^2=1.00$ in HapMap CEPH) with the CAPS SNP rs2033178 (45). Our block 3 haplotype association was mainly driven by the htSNP rs2373722 which was null after correcting for multiple testing ($P_{adj}=0.14$). Neither the SNP rs474695 identified in our study nor any equivalent proxy was included in the CAPS study.

Although \textit{IGF1} SNP rs4764695 was significantly associated with prostate cancer risk and the main effect was slightly greater for high tumor grade (high grade OR=1.43; low grade OR=1.20; $P_{het}=0.065$), the association with IGF-I blood levels was null. Among the 6 BPC3 studies with IGF blood levels in this report, elevated IGF-I levels were significantly associated with higher PCa risk ($P_{trend}=0.02$); the comparison between the highest IGF-I quartile with the lowest quartile yielded a 21% increased risk of PCa ($P=0.02$). In contrast, the association between IGFBP-3 blood levels and PCa was null ($P_{trend}=0.89$). The IGF-I finding in the BPC3 is consistent with results of a recent meta-analysis of 42 studies (OR=1.21; 95% CI: 1.07, 1.36) (64). In the same meta-analysis, data from 29 studies showed a significant inverse association for IGFBP-3 (OR 0.88; 95% CI, 0.79–0.98) with substantial heterogeneity; the inverse association of IGFBP-3 with PCa risk was seen in retrospective studies, but not prospective studies (64).

Several reasons may explain the lack of association between rs4764695 and IGF-I blood levels. First, the IGF-I measurements reflect systemic levels measured at a single time point prior to PCa diagnosis rather than tissue specific levels. Free IGF-I blood levels, unavailable across these studies, may be a more biologically relevant measure and impacted by rs4764695. Furthermore, the complexity of the IGF pathway is likely not entirely captured in these simple associations and a more complete pathway analysis is warranted. Since the rs4764695 marker lies ~34kb downstream of \textit{IGF1}, this marker may be involved in another pathway entirely. In addition, measurement error for both the genotypes and IGF-I blood levels would lead to non-differential misclassification and a bias towards the null. Lastly, the result for rs4764695 may be spurious, although we have taken steps by setting a stringent alpha level (0.01) and correcting for multiple comparisons.

We identified several \textit{IGFBP3} SNPs strongly associated with IGFBP-3 blood levels among Caucasians but not associated with PCa risk. Given a considerable amount of LD exists within the 5’ region (Figure 2) of
IGFBP3, we tested the independent effect of six correlated SNPs using a multi-marker model and determined that the most significant SNP was rs2854746, a nonsynonymous polymorphism in exon 1 (Gly32Ala). This is in contrast to the promoter polymorphism rs2854744 (A-202C) that has been extensively reported to be associated with IGFBP-3 levels (39, 41, 43, 47-54). This observation has been alluded to in two previous reports (50, 53), but could not be substantiated due to the limited sample sizes and strong LD between these two markers among Caucasians ($r^2=0.85$ among Caucasian PLCO controls). Although each minor allele was associated with 6.3% higher IGFBP-3 blood levels on average, rs2854746 explains only 3.6% of the variation. The lack of an association between rs2854746 and PCa risk, whereas a strong influence on IGFBP3 blood levels exists, supports a Mendelian randomization argument for no etiological effect of IGFBP-3 on incident risk of PCa (66, 67). However caution is needed since confounding or pleiotropic effects would negate this argument (68).

The major strength of this study is the utilization of a large cohort consortium and a comprehensive approach to examine the genetic variation across three genes in the IGF pathway, a strong candidate in prostate carcinogenesis. Specifically, we have the ability to look at the effects of SNPs on pre-diagnostic blood levels as well as risk in the same set of subjects. Although our data in Caucasians limits the generalizability to other ethnic/racial groups, our large sample size allows us to present the overall risk estimates using a 99% CI, reducing the chance of both false-positive and false-negative results. In addition, we reduced the probability of a spurious association due to multiple hypothesis testing by applying a correction across all models, traits, and genetic markers. The IGF pathway is a complex system and we have limited our study to three primary genes. Additional IGF pathway genes need to be investigated and a more comprehensive pathway analysis would be necessary.

While several genome-wide association scans (GWAS) have recently identified multiple susceptibility loci for PCa risk (69-77), only recently has an IGF variant (IGF2-rs7127900) been among them. The Cancer Genetic Markers of Susceptibility (CGEMS) GWAS is the only publicly available database to compare our IGF1 rs4764695 finding. Although rs4764695 was not present on the CGEMS platform, a proxy rs1980236 showed a similar effect, although not statistically significant ($P=0.57$). However, the PLCO study was the first
stage of the CGEMS GWAS and over 70% of this sample is included in this analysis where the observed effect for rs4764695 was 1.11 ($P=0.15$). The initial stages of the GWAS studies include less than 2,000 PCa cases and have limited power, 67%, to detect an OR of 1.20 for a MAF of 0.50. In contrast, our large consortium study allows for focused and comprehensive evaluation of candidate genes among over 6,000 PCa cases, having the necessary power, 99%, to identify an OR of 1.20 for a MAF of 0.50. This is evidenced by multiple GWAS studies identifying common as well as different risk loci. Currently, only 6% of PCa genetic variation is explained by the known loci identified in GWAS studies (78). The missing heritability, aka ‘dark matter’, may reside undetected in the “lower Manhattan” plots and represented by multiple variants (79, 80). For example, CGEMS replicated several variants, $JAZF1$ (GWAS rank 24,407; $P=0.04$) and $MSMB$ (GWAS rank 24,223; $P=0.04$), with little evidence from the first stage. Furthermore, both in vivo (1-3) and in vitro (4-6) studies have demonstrated that the IGF pathway plays a role in tumorigensis, thus making it a strong biological candidate.

In conclusion, a significant association between prostate cancer risk and an $IGF1$ SNP, rs4764695, was identified among Caucasians. Although this is a novel finding, the evidence is still preliminary and further confirmation is needed. The estimated population attributable risk for homozygotic variant carriers is ~5% due to the high frequency of the minor allele ($G=49\%$). This variant could be of greater importance due to the potential for a stronger association with high tumor grade PCa. The association between rs4764695 and cancer is exclusive with prostate cancer as the association with breast cancer risk and this marker has been reported as null in the NCI BPC3 study (54). Furthermore, we provide strong evidence for a novel association between IGFBP-3 blood levels and a nonsynonymous $IGFBP3$ marker in contrast to the previously reported $IGFBP3$ promoter polymorphism. Additional studies, such as fine mapping to determine the causal variant in $IGF1$ and the examination of additional genes in the IGF axis are needed. In summary, preliminary evidence implicates common genetic variation in the $IGF1$ locus with PCa risk.
Materials and Methods

Study Population

The BPC3 and member cohorts have been described in detail elsewhere (81). In brief, the consortium combines resources from seven well-established cohort studies: the American Cancer Society Cancer Prevention Study II (CPS-II) (82), the Alpha-Tocopherol, Beta-Carotene Cancer Prevention (ATBC) Study (83), the European Prospective Investigation into Cancer and Nutrition Cohort (EPIC – comprised of cohorts from Denmark, Great Britain, Germany, Greece, Italy, the Netherlands, Spain, and Sweden) (84), the Health Professionals Follow-up Study (HPFS) (85), the Hawaii/Los Angeles MEC (86), the Physicians’ Health Study (PHS) (16), and the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial (PLCO) (87). These cohorts collectively include over 248,000 men with a blood sample.

The current study was restricted to individuals who self-reported as Caucasian and consists of 6,012 prospective PCa cases and 6,641 controls. Cases from other ethnic groups were contributed mostly from the MEC and had been reported on previously (53, 61, 62); we analyzed the data for Caucasians with and without the contribution from the MEC (457 cases and 452 controls) to assess the impact on the overall results. Prospective PCa cases were identified through population-based cancer registries or self-reports confirmed by medical records, including pathology reports. The BPC3 data for PCa consists of a series of matched nested case-control studies within each cohort; controls were matched to cases on a number of potential confounding factors, including age, country, and region of recruitment. For the current analysis, PCa cases were matched to available controls by age in five year intervals, cohort, and country of residence for EPIC. A written informed consent was obtained from all subjects and each study was approved by the Institutional Review Boards at their respective institutions.

IGF-I and IGFBP-3 Blood Levels

Pre-diagnostic measurements of IGF-I and IGFBP-3 were available for six of the seven BPC3 cohort members (ATBC, EPIC, HPFS, MEC, PHS and PLCO; IGF-I: N=6,076; IGFBP-3: N=6,059) (16, 19-23, 53). Most blood samples in the CPS-II were collected post-diagnosis and therefore were only included in the
genotyping analyses. Samples from three of the studies (ATBC, HPFS and PHS) were measured in the Pollak laboratory and the remaining three studies (EPIC, MEC and PLCO) were measured in the laboratory of the Hormones and Cancer Team at IARC; all used enzyme-linked immunosorbent assays (Diagnostic System Laboratories, Webster, TX).

**SNP Discovery and htSNP Selection**

As previously described by Cheng *et al.*, a multi-stage approach was used to characterize genetic variation across *IGF1*, *IGFBP1*, and *IGFBP3* loci (61, 62). Most of the exons across the three genes were resequenced in 95 advanced PCa cases and a multiethnic panel of 349 controls was utilized to determine the patterns of linkage disequilibrium (LD) encompassing ~20 kilobases (kb) upstream and ~10 kb downstream of each gene. Haplotype-tagging SNPs (htSNPs) for each haplotype block, determined by the confidence interval method of Gabriel *et al.* (88, 89), were chosen based on $R^2_h$, a measure of the correlation between observed and predicted haplotypes based on the htSNP genotypes (90), to select a minimum set of SNPs that would achieve an $R^2_h \geq 0.7$ for all common haplotypes with an estimated frequency of $\geq 5\%$ among Caucasians.

For genetic characterization of *IGF1* (chromosome 12q22-q23), 154 SNPs were evaluated over a 156-kb region (1 SNP every 2.4 kb) in a multiethnic panel of subjects with no history of cancer (61). After removing markers that were monomorphic or had poor genotyping results, a panel of 64 SNPs remained, from which 14 htSNPs were selected to predict the common haplotypes among Caucasians ($R^2_h > 0.85$). Of the 14 htSNPs, 11 SNPs are available in HapMap Phase II CEPH samples and capture $60\%$ of the common genetic variation (MAF $> 5\%$) of *IGF1* with an $r^2 > 0.70$.

For genetic characterization of *IGFBP1* and *IGFBP3* (which are located contiguously in a 35-kb region on chromosome 7p13-p12), 56 SNPs over a 71-kb region were evaluated (1 SNP every 2 kb) in the multiethnic panel (61, 62). Twenty markers were removed due to being monomorphic or having poor genotyping results. The final selection included twelve htSNPs to predict the common haplotypes among Caucasians ($R^2_h > 0.99$), two genic SNPs in *IGFBP3* not part of a haplotype block (rs6670, rs2453839), and two additional *IGFBP3* SNPs (rs2132570, rs2960436). Ten of these 16 SNPs are available in HapMap Phase II CEPH samples and
capture 41% of the common genetic variation (MAF>5%) of IGF1 and IGFBP3 with an $r^2>0.70$.

Genotyping

Genotyping was conducted by five laboratories (University of Southern California, Los Angeles, CA, USA; University of Hawaii, Honolulu, HI, USA; Harvard School of Public Health, Boston, MA, USA; Core Genotyping Facility, National Cancer Institute, Bethesda, MD, USA; and Cambridge University, Cambridge, UK) using a fluorescent 5’endonuclease assay and the ABI-PRISM 7900 for sequence detection (TaqMan; Applied Biosystems, Inc.). Assay information is available at the MEC Genetics website (http://uscnorris.com/mecgenetics/CohortGCKView.aspx). For each assay, the concordance rate was 100% for 102 samples from the SNP500 Cancer project (http://snp500cancer.nci.nih.gov) (91) and inter-laboratory completion and concordance rates were greater than 99%, based on cross-laboratory assessment of 30 SNPs on 94 samples from the Coriell Biorepository (Camden, NJ). The internal quality of genotype data at each genotyping center was assessed by typing 5-10% blinded samples in duplicate or greater (depending on study).

Statistical Analysis

All statistical tests presented are two-sided and were conducted in SAS 9.0 (SAS Institute). Figures and multiple testing corrections were generated in the statistical program R (http://cran.r-project.org/). To account for multiple hypothesis testing, we applied the method implemented in PACT, a flexible and efficient approach that adjusts for correlation between multiple traits, genetic markers, and models (92). PACT utilizes less computational time while maintaining the accuracy of permutation or simulation-based tests. Since multiple models and traits can be considered, the p-value corrections were computed simultaneously for blood levels and PCa risk across all 30 IGF1 and IGFBP1/IGFBP3 markers. A test of significance was set at the 0.01 level to minimize the chance of both false positive and false negative results (93, 94). We present the uncorrected p-values ($P_{uncorr}$) for association in the figures and tables. A corrected p-value ($P_{corr}$) is presented for significant SNPs correcting for multiple comparisons across all traits, genetic markers, and statistical models among the 30 IGF genetic markers analyzed.
Genetic Determinants of Blood Levels

We identified cohort- and assay batch-specific statistical outliers based on the generalized extreme studentized deviate many-outlier detection approach (95) setting alpha to 0.05 for both IGF-I and IGFBP-3 blood levels. Based on this, we excluded 23 IGF-I samples (n=14 cases and 9 controls) and 43 IGFBP-3 samples (n=19 cases and 24 controls). The IGF-I and IGFBP-3 blood levels were log-transformed to provide approximate normal distributions. The geometric mean and 95% CI according to haplotypes or SNPs were calculated using linear regression analysis adjusting for age at blood draw, assay batch, cohort (including country for EPIC) and case-control status. Haplotype frequencies and subject-specific expected haplotype indicators were calculated among cases and controls combined by cohort (and country within EPIC) and then combined for the haplotype analyses. A global haplotype test was performed for each haplotype block by using an F test to compare the sum of the squared residuals for a full model (all haplotypes within a block) and a nested model (without haplotypes within a block). A multi-SNP model was utilized by including all statistically significant SNPs identified from the univariate analysis to assess independent SNP effects within a gene.

Genetic Determinants of Prostate Cancer

The statistical methods used have been described previously (93, 96). In brief, we used conditional logistic regression to estimate ORs and 99% CI for disease associated with genetic markers (SNP or haplotype). The matching factors in the conditional logistic regression were age (in five year intervals), cohort, and country within EPIC. We estimated the genotypic ORs for disease by using the most common genotype as the referent group for the SNP analyses. We estimated haplotype-specific ORs using an expectation-substitution approach to account for haplotype uncertainty given unphased genotype data (97, 98). To test the global null hypothesis of no association between IGF genetic variation and risk of PCa, we used a likelihood ratio test comparing a model with additive effects for each common haplotype (treating the most common haplotype as the referent) to the intercept-only model. We considered haplotypes with greater than 5% frequency in at least one cohort to be "common". All other haplotypes were excluded.
When assessing genetic effects (SNPs or haplotype) we tested for heterogeneity across cohort and several potential effect modifiers. We used a likelihood ratio test (LRT) by including an interaction term between the genetic effect (SNP or haplotype) and variable of interest in comparison to the null model. The tests for heterogeneity by cohort were not statistically significant ($P>0.01$), therefore we present the pooled results. We assessed for the presence of effect modification by family history (at least one first-degree relative or two second-degree relatives diagnosed with PCa), age at diagnosis ($<65, \geq 65$), body mass index (BMI) at baseline ($<25, 25-<30, >30$), and height (tertiles and quartiles, cohort specific cut-points among controls were used). Finally, we evaluated the association of gene variants with advanced stage PCa (stage C or D) and high-grade PCa (Gleason score $\geq 8$ or poorly differentiated).

**Acknowledgements**

This work was funded by NCI grants U01 CA098216 (EPIC) U01 CA098233 (Harvard), U01 CA098758 (MEC) and U01 CA098710 (ACS). We would like to thank Hardeep Ranu and Patrice Soule of the Dana-Farber/Harvard Cancer Center High Throughput Polymorphism Detection Core. FRS was supported by the National Research Service Award Training Program in Cancer Epidemiology (T32 CA009001-32). The Physicians’ Health Study is supported by the National Cancer Institute (CA-34944, CA-40360, and CA-097193) and the National Heart, Lung, and Blood Institute (HL-26490 and HL-34595). The Multiethnic Cohort is supported by CA54281 and CA63464. The ATBC Study has been supported by U.S. Public Health Service contracts N01-CN-45165, N01-RC-45035, and N01-RC-37004 from the National Cancer Institute, Department of Health and Human Services.
References


Legends to Figures

Figure 1. *IGF1* linkage disequilibrium and SNP associations with IGF-I blood levels and prostate cancer risk among Caucasians in the BPC3. The upper panel is a plot of the $-\log_{10}$ p-values for *IGF1* htSNP associations with IGF-I blood levels (green triangles; N=5,684) and prostate cancer risk (red circles; 6,012 PCa cases/6,641 controls). The direction of the green triangles indicates the direction of effect between the minor allele and IGF-I blood levels. The *IGF1* locus is illustrated on the x-axis with the exons colored blue and the introns yellow. The arrow indicates the 5'->3' direction. The p-values are plotted according to genetic position. The bottom panel contains a linkage disequilibrium (LD) plot for the 14 *IGF1* htSNPs among PLCO Caucasian controls. The $r^2$ values are used for LD. The black triangles indicate the haplotype blocks. Corrected p-values for marginally associated SNPs with prostate cancer risk: rs2373722=0.14; rs4764695=8.7x10^{-3}.

Figure 2. *IGFBP1* and *IGFBP3* linkage disequilibrium and SNP associations with IGFBP-3 blood levels and prostate cancer risk among Caucasians in the BPC3. The upper panel is a plot of the $-\log_{10}$ p-values for *IGFBP1* and *IGFBP3* SNP associations with IGFBP-3 blood levels (green triangles; N=5,667) and prostate cancer risk (red circles; 6,012 PCa cases/6,641 controls). The direction of the green triangles indicates the direction of effect between the minor allele and IGFBP-3 blood levels. The *IGFBP1* and *IGFBP3* loci are illustrated on the x-axis with the exons colored blue and the introns yellow. The arrow indicates the 5'->3' direction. The p-values are plotted according to genetic position. The bottom panel contains a linkage disequilibrium (LD) plot for the 16 *IGFBP1* and *IGFBP3* SNPs among PLCO Caucasian controls. The $r^2$ values are used for LD. The black triangles indicate the haplotype blocks. Corrected p-values for blood level associations: rs2270628=8.6x10^{-3}; rs3110697=1.1x10^{-16}; rs2854746=8.8x10^{-43}; rs2854744=8.1x10^{-34}; rs2132570=1.8x10^{-4}; rs2960436=4.4x10^{-34}.

Figure 3. Forest plot of risk estimates for rs4764695 with prostate cancer among Caucasians in the seven cohorts and combined. The left side list of the figure lists the cohort member, the case/control genotype distribution, and the minor allele frequency (MAF). The blue squares and connected lines represent the cohort-specific genotype risk estimates and the 99% CI. The genotype summary risk estimates and 99% CIs are represented by the red diamonds. Corrected p-value=8.7x10^{-3}.
<table>
<thead>
<tr>
<th>Characteristics of Caucasians in the Breast &amp; Prostate Cancer Cohort Consortium (BPC3) by cohort and overall.</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPS-II</td>
</tr>
<tr>
<td>Cases</td>
</tr>
<tr>
<td>Number</td>
</tr>
<tr>
<td>Age at diagnosis (mean), yrs</td>
</tr>
<tr>
<td>BMI† (mean), kg/m²</td>
</tr>
<tr>
<td>Height (mean), cm</td>
</tr>
<tr>
<td>Family history available‡ (n)</td>
</tr>
<tr>
<td>Family history (% yes)</td>
</tr>
</tbody>
</table>

For cases:

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Stage info available (n)</td>
<td>1,129</td>
<td>596</td>
<td>405</td>
<td>570</td>
<td>437</td>
<td>668</td>
<td>0</td>
<td>3,805</td>
</tr>
<tr>
<td>Stage (% ≥ C)</td>
<td>11</td>
<td>31</td>
<td>17</td>
<td>15</td>
<td>14</td>
<td>31</td>
<td>n/a</td>
<td>19</td>
</tr>
<tr>
<td>Gleason score available</td>
<td>999</td>
<td>583</td>
<td>97</td>
<td>583</td>
<td>445</td>
<td>666</td>
<td>0</td>
<td>3,373</td>
</tr>
<tr>
<td>Gleason score (% ≥ 8)</td>
<td>11</td>
<td>26</td>
<td>16</td>
<td>9</td>
<td>26</td>
<td>13</td>
<td>n/a</td>
<td>16</td>
</tr>
</tbody>
</table>

†BMI: Body Mass Index; ‡Family history: at least one first-degree relative or two second-degree relatives diagnosed with prostate cancer
Table 2. Stratified associations between *IGFBP3* SNPs and IGFBP-3 blood levels.

<table>
<thead>
<tr>
<th></th>
<th>rs2854746</th>
<th>rs2854744</th>
<th>P^‡</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GG</td>
<td>GC</td>
<td>CC</td>
</tr>
<tr>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Total</td>
<td>1,918</td>
<td>2,597</td>
<td>875</td>
</tr>
<tr>
<td>N Mean (95% CI)</td>
<td>3,046 (2,975-3,119)</td>
<td>3,263 (3,188-3,339)</td>
<td>3,442 (3,353-3,533)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>rs2854746</th>
<th>rs2854744</th>
<th>P^‡</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1,573</td>
<td>301</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>2,343</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>875</td>
<td>858</td>
</tr>
</tbody>
</table>

†Mean IGFBP-3 blood levels (nanograms per microliter) and 95% confidence intervals
‡SNP stratified p-value from linear regression adjusting for age at blood draw, assay batch, prostate cancer, cohort, and country
Table 3. Risk estimates for *IGF1* SNP rs4764695 and prostate cancer aggressiveness among Caucasians in the NCI Breast & Prostate Cancer Cohort Consortium. Corrected p-value for overall PCa risk=8.7x10^{-3}.

<table>
<thead>
<tr>
<th>SNP</th>
<th>Genotype</th>
<th>Case (%)</th>
<th>Control (%)</th>
<th>OR† (99% CI‡)</th>
<th>P§</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs4764695</td>
<td>AA</td>
<td>1,396 (24.0)</td>
<td>1,612 (25.3)</td>
<td>Ref.</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>AG</td>
<td>2,800 (48.1)</td>
<td>3,203 (50.2)</td>
<td>1.01 (0.9-1.14)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GG</td>
<td>1,631 (28.0)</td>
<td>1,561 (24.5)</td>
<td>1.20 (1.06-1.37)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Case (%)</th>
<th>Control (%)</th>
<th>OR† (99% CI‡)</th>
<th>P§</th>
</tr>
</thead>
<tbody>
<tr>
<td>High Grade (≥8)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AA</td>
<td>123 (23.6)</td>
<td>1,612 (25.3)</td>
<td>Ref.</td>
<td>0.001</td>
</tr>
<tr>
<td>AG</td>
<td>234 (44.9)</td>
<td>3,203 (50.2)</td>
<td>0.97 (0.71-1.31)</td>
<td></td>
</tr>
<tr>
<td>GG</td>
<td>164 (31.5)</td>
<td>1,561 (24.5)</td>
<td>1.43 (1.02-1.99)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Case (%)</th>
<th>Control (%)</th>
<th>OR† (99% CI‡)</th>
<th>P§</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low Grade (&lt;8)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AA</td>
<td>919 (21.8)</td>
<td>1,951 (22.6)</td>
<td>Ref.</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>AG</td>
<td>1,935 (46.0)</td>
<td>4,255 (49.2)</td>
<td>1.00 (0.87-1.14)</td>
<td></td>
</tr>
<tr>
<td>GG</td>
<td>1,352 (32.1)</td>
<td>2,442 (28.2)</td>
<td>1.20 (1.04-1.40)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Case (%)</th>
<th>Control (%)</th>
<th>OR† (99% CI‡)</th>
<th>P§</th>
</tr>
</thead>
<tbody>
<tr>
<td>High Stage (C or D)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AA</td>
<td>1,612 (25.3)</td>
<td>Ref.</td>
<td>0.107</td>
<td></td>
</tr>
<tr>
<td>AG</td>
<td>3,203 (50.2)</td>
<td>1.01 (0.78-1.31)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GG</td>
<td>1,561 (24.5)</td>
<td>1.22 (0.91-1.64)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Case (%)</th>
<th>Control (%)</th>
<th>OR† (99% CI‡)</th>
<th>P§</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low Stage (A or B)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AA</td>
<td>736 (24.4)</td>
<td>1,612 (25.3)</td>
<td>Ref.</td>
<td>0.002</td>
</tr>
<tr>
<td>AG</td>
<td>1,449 (48.0)</td>
<td>3,203 (50.2)</td>
<td>1.05 (0.90-1.22)</td>
<td></td>
</tr>
<tr>
<td>GG</td>
<td>834 (27.6)</td>
<td>1,561 (24.5)</td>
<td>1.24 (1.04-1.47)</td>
<td></td>
</tr>
</tbody>
</table>

†Conditional logistic regression models adjusted for age in 5-year intervals, cohort, and country
‡99% confidence intervals
§P-value for association from 2 df likelihood ratio test
Figure 1

IGF1 (12q22-q23)

- \(-\log_{10}(P\text{-value})\)

\(P=0.01\)

Physical Distance: 139.7 kb
LD Map Type: r\(^2\)

IGF1 Blood Levels
\(\bullet\) PCa Risk
Figure 2

IGFBP1 & IGFBP3 (7p13–p12)

Physical Distance: 68.4 kb
LD Map Type: r-square

△ IGFBP–3 Blood Levels
● PCa Risk
Figure 3

<table>
<thead>
<tr>
<th>Cohort</th>
<th>Case/Control</th>
<th>MAF</th>
<th>rs4764695</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPSII</td>
<td>298/319</td>
<td>0.48</td>
<td>Ref.</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>536/554</td>
<td></td>
<td>AG</td>
<td></td>
</tr>
<tr>
<td></td>
<td>301/262</td>
<td></td>
<td>GG</td>
<td></td>
</tr>
<tr>
<td>ATBC</td>
<td>200/206</td>
<td>0.53</td>
<td>Ref.</td>
<td>0.32</td>
</tr>
<tr>
<td></td>
<td>479/489</td>
<td></td>
<td>AG</td>
<td></td>
</tr>
<tr>
<td></td>
<td>298/201</td>
<td></td>
<td>GG</td>
<td></td>
</tr>
<tr>
<td>EPIC</td>
<td>151/230</td>
<td>0.5</td>
<td>Ref.</td>
<td>0.78</td>
</tr>
<tr>
<td></td>
<td>322/481</td>
<td></td>
<td>AG</td>
<td></td>
</tr>
<tr>
<td></td>
<td>163/229</td>
<td></td>
<td>GG</td>
<td></td>
</tr>
<tr>
<td>HPFS</td>
<td>151/162</td>
<td>0.49</td>
<td>Ref.</td>
<td>0.007</td>
</tr>
<tr>
<td></td>
<td>297/333</td>
<td></td>
<td>AG</td>
<td></td>
</tr>
<tr>
<td></td>
<td>194/144</td>
<td></td>
<td>GG</td>
<td></td>
</tr>
<tr>
<td>MEC</td>
<td>126/119</td>
<td>0.48</td>
<td>Ref.</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td>210/230</td>
<td></td>
<td>AG</td>
<td></td>
</tr>
<tr>
<td></td>
<td>120/99</td>
<td></td>
<td>GG</td>
<td></td>
</tr>
<tr>
<td>PHS</td>
<td>195/271</td>
<td>0.48</td>
<td>Ref.</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>398/449</td>
<td></td>
<td>AG</td>
<td></td>
</tr>
<tr>
<td></td>
<td>223/238</td>
<td></td>
<td>GG</td>
<td></td>
</tr>
<tr>
<td>PLCO</td>
<td>275/305</td>
<td>0.51</td>
<td>Ref.</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>558/667</td>
<td></td>
<td>AG</td>
<td></td>
</tr>
<tr>
<td></td>
<td>332/328</td>
<td></td>
<td>GG</td>
<td></td>
</tr>
<tr>
<td>NCI Breast &amp;</td>
<td>1,396/1,612</td>
<td>0.5</td>
<td>Ref.</td>
<td>0.00011</td>
</tr>
<tr>
<td>Prostate Cancer</td>
<td>2,800/3,203</td>
<td></td>
<td>AG</td>
<td></td>
</tr>
<tr>
<td>Cohort Consortium</td>
<td>1,631/1,561</td>
<td></td>
<td>GG</td>
<td></td>
</tr>
</tbody>
</table>

OR (99% CI)
Abbreviations

ATBC: Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study
Ala: Alanine
BMI: Body mass index
BPC3: Breast & Prostate Cancer Cohort Consortium
CAPS: Caner Prostate in Sweden
CGEMS: Cancer Genetic Markers of Susceptibility
CI: Confidence interval
CPS-II: American Cancer Society Cancer Prevention Study
EPIC: European Prospective Investigation into Cancer and Nutrition Cohort
Gly: Glycine
GWAS: Genome-wide association scan
HPFS: Health Professionals Follow-up Study
htSNP: Haplotype-tagging single nucleotide polymorphism
IGF: Insulin-like growth factor
IGFBP: Insulin-like growth factor binding protein
Kb: Kilobase
LD: Linkage disequilibrium
LRT: Likelihood ratio test
MAF: Minor allele frequency
MEC: Multiethnic Cohort
MRBS: mRNA by SNP Browser
NCI: National Cancer Institute
OR: Odds ratio
PCa: Prostate cancer
PHS: Physicians’ Health Study
PLCO: Prostate, Lung, Colorectal, Ovarian Cancer Screening Trial
SNP: Single nucleotide polymorphism