## SUPPLEMENTARY MATERIAL for

# Identification of a novel prostate cancer susceptibility locus on chromosome 8q24

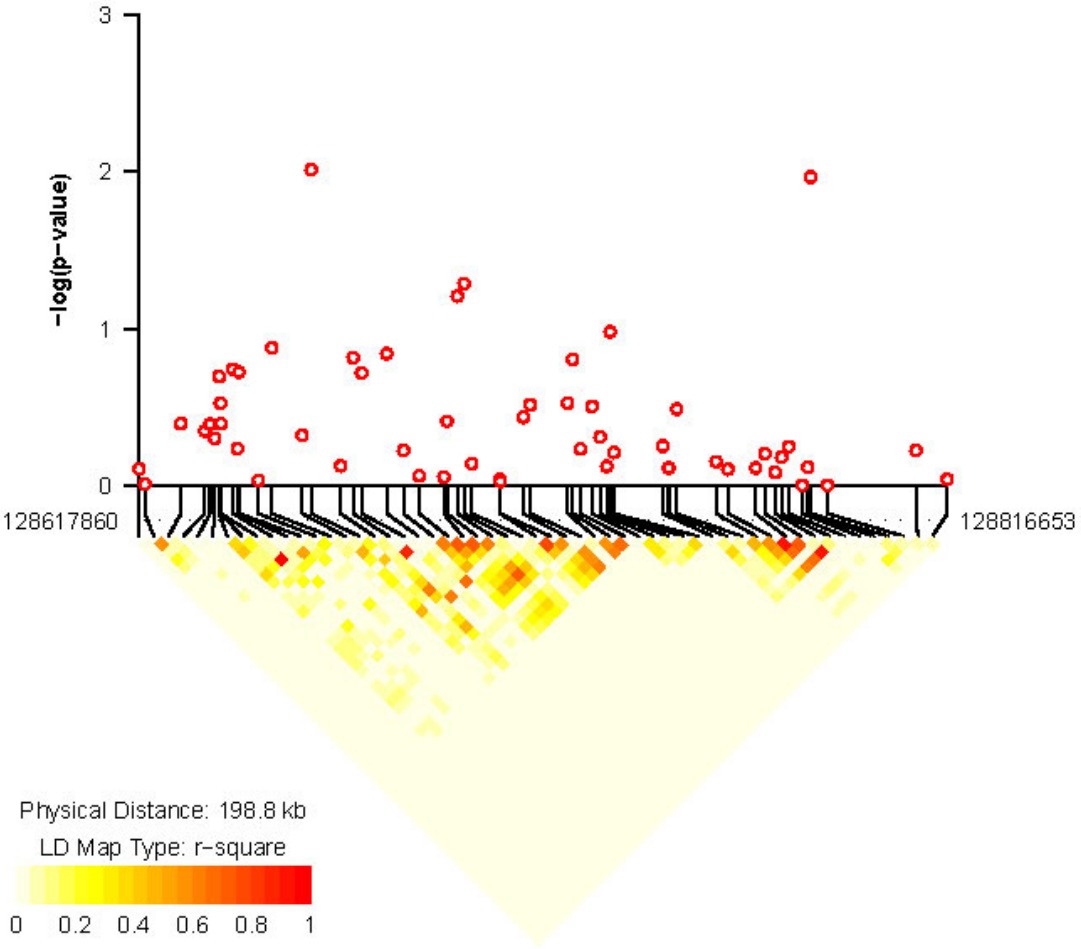
Meredith Yeager[1,2+], Nilanjan Chatterjee[2], Julia Ciampa[2], Kevin B. Jacobs[3], Jesus Gonzalez Bosquet[2], Richard B. Hayes[2], Peter Kraft[4], Sholom Wacholder[2], Nick Orr[2], Sonja Berndt[2], Kai Yu[2], Amy Hutchinson[1,2], Zhaoming Wang[1,2], Laufey Amundadottir[2], Heather Spencer Feigelson[5], Michael J. Thun[5], W. Ryan Diver[5], Demetrius Albanes[2], Jarmo Virtamo[6], Stephanie Weinstein[2], Fredrick R. Schumacher[4,7], Geraldine Cancel-Tassin[8], Olivier Cussenot[8], Antoine Valeri[8], Gerald L Andriole[9], E. David Crawford[10], Christopher A. Haiman[11], Brian Henderson[11], Laurence Kolonel[12], Loic Le Marchand[12], Afshan Siddiq[13], Elio Riboli[13], Timothy J. Key[14], Rudolf Kaaks[15], William Isaacs[16], Sarah Isaacs[16], Kathleen E. Wiley[16], Henrik Gronberg[17], Fredrik Wiklund[17], Pär Stattin[18], Jianfeng Xu[19], S. Lilly Zheng[19], Jielin Sun[19], Lars J. Vatten[20], Kristian Hveem[20], Merethe Kumle[21], Margaret Tucker[2], Daniela S. Gerhard[22], Robert N. Hoover[2], Joseph F. Fraumeni, Jr[2], David J. Hunter[4,23], Gilles Thomas[2], Stephen J. Chanock[2]

[1] Core Genotyping Facility, Advanced Technology Program, SAIC-Frederick, Inc., NCI-Frederick, Frederick, MD 21702.
[2] Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Heath, Department of Health and Human Services, Bethesda, Maryland 20892, USA.
[3] Bioinformed Consulting Services, Gaithersburg, Maryland 20877, USA.
[4] Program in Molecular and Genetic Epidemiology, Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts 02115, USA.
[5]Department of Epidemiology and Surveillance Research, American Cancer Society, Atlanta, GA
[6] Department of Chronic Disease Prevention, National Institute for Health and Welfare, Helsinki, FIN-00300, Finland.
[7] Department of Nutrition, Harvard School of Public Health, Boston, Massachusetts 02115, USA.
[8] Centre de Recherche pour les Pathologies Prostatiques (CeRePP), Hôpital Tenon, Assistance Publique-Hôpitaux de Paris, 75970 Paris, France.
[9] Division of Urologic Surgery, Washington University School of Medicine, St. Louis, Missouri 63108, USA.
[10] Department of Surgery, University of Colorado at Denver and Health Sciences Center, Denver, CO, USA.
[11] Department of Preventive Medicine, Keck School of Medicine, Norris Comprehensive Cancer Center, University of Southern California, Los Angeles, CA
[12] Epidemiology Program, Cancer Research Center, University of Hawaii, Honolulu, HI
[13] Division of Epidemiology, Public Health and Primary Care, Imperial College, London, England
[14] Cancer Epidemiology Unit, University of Oxford, Oxford, United Kingdom
[15] Division of Clinical Epidemiology, German Cancer Research Centre (DKFZ), Heidelberg, Germany
[16] Johns Hopkins Hospital, Baltimore, MD 21287
[17] Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm
[18] Department of Surgical and Perioperative Sciences, Urology and Andrology, Umeå University, Sweden
[19] Center for Cancer Genomics, Wake Forest University School of Medicine, Winston-Salem, NC 27157
[20] Department of Public Health and General Practice, Norwegian University of Science and Technology, NO-7489 Trondheim, Norway
[21] Institute of Community Medicine, University of Tromsø, Norway
[22] Office of Cancer Genomics, NCI, NIH, DHHS Bethesda, MD 20892
[23] Channing Laboratory, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA

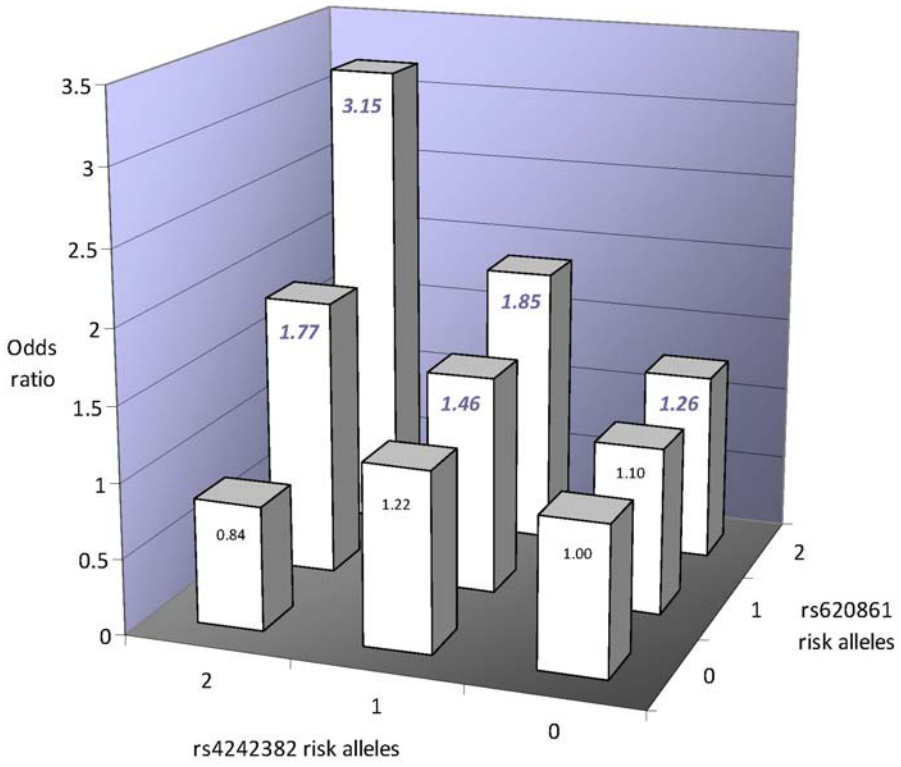
[+]Correspondence to:

Meredith Yeager, Ph.D.
Core Genotyping Facility
Advanced Technology Center
National Cancer Institute
8717 Grovemont Circle
Gaithersburg, MD 20877 (FedEx)
Bethesda, MD 20892 (US Mail)
Phone: 301-435-7613
Fax: 301-443-7088

**Supplementary Figure 1.** Linkage disequilibrium (LD) and Association Results for 55 SNPs Genotyped Over ~200kb of Chromosome 8q24



**Supplementary Figure 2. Joint Odds-Ratio Adjusted For Study for rs4242382 and rs620861 Risk Alleles** Italics indicate ORs are significant at the 5% significance level. p-value for 1 df test for multiplicative interaction = $2.2 \times 10^{-3}$.



**Supplementary Table 1. Detailed association results for rs620861 and rs7841060**

## rs620861

### Adjusted genotype test, dichotomous phenotype

| SUBSET | Rank | MAF | Subjects | score $X^2$ | score p-value | df | het OR | het OR 95% CI | | hom OR | hom OR 95% CI | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| STAGE1-3 | 37 | 0.372\|0.338 | 9108\|10256 | 45.58 | 1.3E-10 | 2 | 1.17 | (1.1 - | 1.24) | 1.33 | (1.21 - | 1.45) |
| STAGE1 | 1338 | 0.383\|0.356 | 926\|971 | 4.43 | 1.1E-01 | 2 | 1.22 | (1.01 - | 1.49) | 1.21 | (.91 - | 1.61) |
| PLCO | 1338 | 0.383\|0.356 | 926\|971 | 4.43 | 1.1E-01 | 2 | 1.22 | (1.01 - | 1.49) | 1.21 | (.91 - | 1.61) |
| STAGE2 | 45 | 0.371\|0.341 | 4044\|4127 | 17.08 | 2.0E-04 | 2 | 1.12 | (1.02 - | 1.22) | 1.33 | (1.16 - | 1.53) |
| CPSII | 125 | 0.383\|0.343 | 1639\|1632 | 10.98 | 4.1E-03 | 2 | 1.16 | (1.01 - | 1.35) | 1.42 | (1.14 - | 1.77) |
| ATBC | 139 | 0.353\|0.307 | 866\|905 | 8.71 | 1.3E-02 | 2 | 1.25 | (1.02 - | 1.52) | 1.54 | (1.1 - | 2.14) |
| HPFS | 1589 | 0.357\|0.380 | 589\|594 | 2.30 | 3.2E-01 | 2 | 0.83 | (.65 - | 1.06) | 0.88 | (.62 - | 1.25) |
| FPCC | 487 | 0.377\|0.343 | 950\|996 | 5.29 | 7.1E-02 | 2 | 1.12 | (.93 - | 1.36) | 1.39 | (1.04 - | 1.86) |
| STAGE3 | 26 | 0.370\|0.332 | 4138\|5158 | 27.27 | 1.2E-06 | 2 | 1.21 | (1.11 - | 1.33) | 1.35 | (1.18 - | 1.55) |
| CONOR | 1297 | 0.359\|0.328 | 659\|605 | 3.05 | 2.2E-01 | 2 | 1.21 | (.96 - | 1.54) | 1.24 | (.87 - | 1.76) |
| JHU | 1201 | 0.377\|0.345 | 451\|988 | 3.06 | 2.2E-01 | 2 | 1.14 | (.9 - | 1.45) | 1.36 | (.95 - | 1.95) |
| SWEDEN | 123 | 0.365\|0.326 | 1356\|2208 | 8.25 | 1.6E-02 | 2 | 1.17 | (1.01 - | 1.36) | 1.35 | (1.07 - | 1.69) |
| MEC | 162 | 0.387\|0.341 | 682\|675 | 6.72 | 3.5E-02 | 2 | 1.30 | (1.03 - | 1.63) | 1.42 | (1.01 - | 2.01) |
| EPIC | 253 | 0.368\|0.324 | 990\|682 | 6.86 | 3.2E-02 | 2 | 1.26 | (1.02 - | 1.55) | 1.41 | (1.02 - | 1.95) |

**Adjusted genotype test, trichotomous phenotype**

| SUBSET | Rank | MAF | Subjects | score $X^2$ | score p-value | df | het1 OR | het1 OR 95% CI | | hom1 OR | hom1 OR 95% CI | | het2 OR | het2 OR 95% CI | | hom2 OR | hom2 OR 95% CI | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| STAGE1-3 | 37 | 0.373\|0.335\|0.340 | 8449\|4532\|4603 | 47.04 | 1.5E-09 | 4 | 1.21 | (1.12 - | 1.31) | 1.33 | (1.18 - | 1.5) | 1.13 | (1.04 - | 1.22) | 1.36 | (1.2 - | 1.53) |
| STAGE1 | 2237 | 0.383\|0.351\|0.361 | 926\|425\|546 | 6.14 | 1.9E-01 | 4 | 1.35 | (1.05 - | 1.73) | 1.21 | (.84 - | 1.74) | 1.14 | (.9 - | 1.43) | 1.21 | (.86 - | 1.7) |
| PLCO | 2237 | 0.383\|0.351\|0.361 | 926\|425\|546 | 6.14 | 1.9E-01 | 4 | 1.35 | (1.05 - | 1.73) | 1.21 | (.84 - | 1.74) | 1.14 | (.9 - | 1.43) | 1.21 | (.86 - | 1.7) |
| STAGE2 | 94 | 0.371\|0.341\|0.338 | 4044\|1827\|1937 | 19.25 | 7.0E-04 | 4 | 1.11 | (.98 - | 1.25) | 1.30 | (1.08 - | 1.56) | 1.12 | (1. - | 1.26) | 1.43 | (1.19 - | 1.72) |
| CPSII | 205 | 0.383\|0.345\|0.337 | 1639\|650\|823 | 12.52 | 1.4E-02 | 4 | 1.15 | (.95 - | 1.4) | 1.41 | (1.05 - | 1.89) | 1.17 | (.98 - | 1.4) | 1.54 | (1.17 - | 2.02) |
| ATBC | 138 | 0.353\|0.307\|0.301 | 866\|526\|244 | 12.14 | 1.6E-02 | 4 | 1.30 | (1.03 - | 1.63) | 1.45 | (.99 - | 2.13) | 1.10 | (.82 - | 1.47) | 2.20 | (1.21 - | 4.01) |
| HPFS | 3178 | 0.357\|0.381\|0.372 | 589\|416\|109 | 2.61 | 6.2E-01 | 4 | 0.81 | (.62 - | 1.06) | 0.87 | (.59 - | 1.29) | 0.98 | (.63 - | 1.52) | 0.86 | (.47 - | 1.59) |
| FPCC | 1404 | 0.377\|0.338\|0.345 | 950\|235\|761 | 5.48 | 2.4E-01 | 4 | 1.11 | (.82 - | 1.5) | 1.51 | (.93 - | 2.46) | 1.13 | (.92 - | 1.38) | 1.36 | (1. - | 1.85) |
| STAGE3 | 21 | 0.372\|0.327\|0.337 | 3479\|2280\|2120 | 28.31 | 1.1E-05 | 4 | 1.29 | (1.15 - | 1.45) | 1.39 | (1.16 - | 1.65) | 1.14 | (1.01 - | 1.28) | 1.35 | (1.12 - | 1.62) |
| JHU | 1302 | 0.377\|0.329\|0.361 | 451\|488\|497 | 5.50 | 2.4E-01 | 4 | 1.24 | (.94 - | 1.63) | 1.57 | (1.03 - | 2.41) | 1.05 | (.8 - | 1.38) | 1.18 | (.78 - | 1.79) |
| SWEDEN | 224 | 0.365\|0.326\|0.327 | 1356\|991\|1180 | 9.85 | 4.3E-02 | 4 | 1.23 | (1.03 - | 1.47) | 1.28 | (.97 - | 1.69) | 1.12 | (.95 - | 1.32) | 1.39 | (1.07 - | 1.82) |
| MEC | 145 | 0.387\|0.344\|0.334 | 682\|339\|223 | 11.46 | 2.2E-02 | 4 | 1.48 | (1.12 - | 1.96) | 1.25 | (.84 - | 1.87) | 1.16 | (.84 - | 1.6) | 1.80 | (1.05 - | 3.09) |
| EPIC | 488 | 0.368\|0.315\|0.343 | 990\|462\|220 | 8.28 | 8.2E-02 | 4 | 1.26 | (1. - | 1.6) | 1.58 | (1.08 - | 2.31) | 1.25 | (.91 - | 1.71) | 1.14 | (.72 - | 1.82) |

## rs7841060

### Adjusted genotype test, dichotomous phenotype

| SUBSET | Rank | MAF | Subjects | score $X^2$ | score p-value | df | het OR | het OR 95% CI | hom OR | hom OR 95% CI |
|---|---|---|---|---|---|---|---|---|---|---|
| STAGE1-3 | 32 | 0.211\|0.246 | 9111\|10257 | 59.09 | 1.5E-13 | 2 | 1.19 | (1.12 - 1.26) | 1.52 | (1.33 - 1.74) |
| STAGE1 | 1796 | 0.217\|0.235 | 926\|971 | 3.44 | 1.8E-01 | 2 | 1.02 | (.84 - 1.24) | 1.49 | (.98 - 2.27) |
| PLCO | 1796 | 0.217\|0.235 | 926\|971 | 3.44 | 1.8E-01 | 2 | 1.02 | (.84 - 1.24) | 1.49 | (.98 - 2.27) |
| STAGE2 | 12 | 0.205\|0.239 | 4047\|4127 | 27.07 | 1.3E-06 | 2 | 1.22 | (1.11 - 1.33) | 1.48 | (1.21 - 1.83) |
| CPSII | 13 | 0.194\|0.244 | 1640\|1632 | 24.18 | 5.6E-06 | 2 | 1.36 | (1.17 - 1.57) | 1.75 | (1.26 - 2.43) |
| ATBC | 1824 | 0.212\|0.233 | 866\|903 | 2.31 | 3.2E-01 | 2 | 1.11 | (.91 - 1.36) | 1.31 | (.85 - 2.02) |
| HPFS | 1803 | 0.215\|0.227 | 589\|595 | 2.08 | 3.5E-01 | 2 | 0.99 | (.78 - 1.26) | 1.53 | (.84 - 2.78) |
| FPCC | 391 | 0.211\|0.241 | 952\|997 | 5.75 | 5.6E-02 | 2 | 1.24 | (1.03 - 1.5) | 1.28 | (.85 - 1.92) |
| STAGE3 | 21 | 0.216\|0.254 | 4138\|5159 | 31.15 | 1.7E-07 | 2 | 1.20 | (1.1 - 1.31) | 1.56 | (1.29 - 1.9) |
| CONOR | 124 | 0.204\|0.250 | 661\|605 | 8.15 | 1.7E-02 | 2 | 1.35 | (1.07 - 1.71) | 1.59 | (.93 - 2.74) |
| JHU | 5071 | 0.243\|0.245 | 451\|990 | 0.02 | 9.9E-01 | 2 | 1.01 | (.8 - 1.28) | 1.03 | (.64 - 1.66) |
| SWEDEN | 59 | 0.227\|0.260 | 1354\|2206 | 9.56 | 8.4E-03 | 2 | 1.20 | (1.04 - 1.39) | 1.43 | (1.05 - 1.94) |
| MEC | 1 | 0.183\|0.265 | 682\|676 | 30.78 | 2.1E-07 | 2 | 1.45 | (1.15 - 1.81) | 4.42 | (2.35 - 8.29) |
| EPIC | 1756 | 0.220\|0.238 | 990\|682 | 2.25 | 3.2E-01 | 2 | 1.04 | (.85 - 1.29) | 1.39 | (.9 - 2.14) |

### Adjusted genotype test, trichotomous phenotype

| SUBSET | Rank | MAF | Subjects | score $X^2$ | score p-value | df | het1 OR | het1 OR 95% CI | hom1 OR | hom1 OR 95% CI | het2 OR | het2 OR 95% CI | hom2 OR | hom2 OR 95% CI |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| STAGE1-3 | 34 | 0.212\|0.240\|0.250 | 8450\|4532\|4604 | 51.54 | 1.7E-10 | 4 | 1.16 | (1.08 - 1.26) | 1.35 | (1.14 - 1.6) | 1.18 | (1.09 - 1.28) | 1.63 | (1.38 - 1.92) |
| STAGE1 | 2472 | 0.217\|0.221\|0.246 | 926\|424\|547 | 5.58 | 2.3E-01 | 4 | 0.97 | (.76 - 1.24) | 1.19 | (.69 - 2.06) | 1.07 | (.85 - 1.34) | 1.73 | (1.08 - 2.76) |
| PLCO | 2472 | 0.217\|0.221\|0.246 | 926\|424\|547 | 5.58 | 2.3E-01 | 4 | 0.97 | (.76 - 1.24) | 1.19 | (.69 - 2.06) | 1.07 | (.85 - 1.34) | 1.73 | (1.08 - 2.76) |
| STAGE2 | 53 | 0.205\|0.233\|0.241 | 4047\|1827\|1937 | 25.11 | 4.8E-05 | 4 | 1.22 | (1.08 - 1.37) | 1.29 | (.99 - 1.69) | 1.23 | (1.1 - 1.39) | 1.52 | (1.18 - 1.95) |
| CPSII | 19 | 0.194\|0.246\|0.240 | 1640\|651\|822 | 25.58 | 3.8E-05 | 4 | 1.49 | (1.23 - 1.81) | 1.45 | (.93 - 2.27) | 1.29 | (1.08 - 1.55) | 1.78 | (1.21 - 2.61) |
| ATBC | 2127 | 0.212\|0.211\|0.252 | 866\|524\|244 | 4.23 | 3.8E-01 | 4 | 1.02 | (.81 - 1.29) | 0.91 | (.53 - 1.56) | 1.25 | (.92 - 1.69) | 1.57 | (.86 - 2.89) |
| HPFS | 2684 | 0.215\|0.230\|0.243 | 589\|417\|109 | 3.17 | 5.3E-01 | 4 | 1.02 | (.79 - 1.33) | 1.53 | (.8 - 2.92) | 1.02 | (.66 - 1.58) | 2.08 | (.84 - 5.14) |
| FPCC | 792 | 0.211\|0.251\|0.238 | 952\|235\|762 | 7.03 | 1.3E-01 | 4 | 1.25 | (.92 - 1.7) | 1.64 | (.9 - 3.) | 1.24 | (1.01 - 1.52) | 1.17 | (.75 - 1.83) |
| STAGE3 | 25 | 0.218\|0.249\|0.258 | 3477\|2281\|2120 | 25.04 | 4.9E-05 | 4 | 1.16 | (1.04 - 1.3) | 1.43 | (1.12 - 1.82) | 1.16 | (1.03 - 1.31) | 1.71 | (1.34 - 2.18) |
| JHU | 755 | 0.243\|0.240\|0.249 | 451\|489\|498 | 7.05 | 1.3E-01 | 4 | 1.10 | (.84 - 1.44) | 0.74 | (.41 - 1.34) | 0.92 | (.7 - 1.21) | 1.30 | (.78 - 2.19) |
| SWEDEN | 329 | 0.227\|0.259\|0.260 | 1354\|990\|1179 | 9.12 | 5.8E-02 | 4 | 1.17 | (.98 - 1.4) | 1.43 | (1. - 2.06) | 1.21 | (1.02 - 1.43) | 1.41 | (1. - 2.) |
| MEC | 1 | 0.183\|0.260\|0.278 | 682\|340\|223 | 31.71 | 2.2E-06 | 4 | 1.38 | (1.05 - 1.82) | 4.40 | (2.19 - 8.83) | 1.50 | (1.09 - 2.07) | 5.28 | (2.51 - 11.08) |
| EPIC | 2004 | 0.220\|0.229\|0.255 | 990\|462\|220 | 4.19 | 3.8E-01 | 4 | 1.01 | (.79 - 1.28) | 1.20 | (.73 - 1.98) | 1.13 | (.82 - 1.55) | 1.83 | (1.01 - 3.32) |

**Supplementary Table 2.** Pairwise estimates of $r^2$ A) per geographic location1 and B) per study for the most highly significant SNPs in prostate regions 1 - 4[1].

A.

| | | | | **EUROPE** | **SCAND** | **USA** |
|---|---|---|---|---|---|---|
| SNP1 | region | SNP2 | region | (1713) | (3119) | (4289) |
| rs620861 | 4 | rs7841060 | 2 | 0.00460 | 0.01124 | 0.00002 |
| rs620861 | 4 | rs4242382 | 1 | 0.00357 | 0.00050 | 0.00170 |
| rs620861 | 4 | rs6983267 | 3 | 0.00528 | 0.01453 | 0.00982 |
| rs4242382 | 1 | rs6983267 | 3 | 0.00500 | 0.00001 | 0.00108 |
| rs4242382 | 1 | rs7841060 | 2 | 0.00330 | 0.00041 | 0.00071 |
| rs6983267 | 3 | rs7841060 | 2 | 0.00009 | 0.00479 | 0.00324 |

[1] EUROPE = FPCC and non-Scandinavian countries represented in the EPIC study; SCAN = ATBC, CONOR, SWEDEN, and Scandinavian countries represented in the EPIC study; USA = CPSII, HPFS, JHU, PLCO

B.

| | | | | **ATBC** | **CONOR** | **CPSII** | **EPIC** | **FPCC** | **HPFS** | **JHU** | **MEC** | **PLCO** | **SWEDEN** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| SNP1 | region | SNP2 | region | (868) | (662) | (1640) | (990) | (952) | (589) | (451) | (682) | (927) | (1362) |
| rs620861 | 4 | rs7841060 | 2 | 0.01250 | 0.01255 | 0.00035 | 0.00511 | 0.00663 | 0.00048 | 0.00004 | 0.00010 | 0.00000 | 0.00933 |
| rs620861 | 4 | rs6983267 | 1 | 0.02312 | 0.00538 | 0.00687 | 0.00634 | 0.00308 | 0.01698 | 0.01016 | 0.00852 | 0.01246 | 0.01867 |
| rs620861 | 4 | rs4242382 | 3 | 0.00136 | 0.00141 | 0.00272 | 0.01113 | 0.00118 | 0.00015 | 0.00057 | 0.00512 | 0.00066 | 0.00110 |
| rs4242382 | 1 | rs6983267 | 3 | 0.00570 | 0.00796 | 0.00347 | 0.00119 | 0.00537 | 0.00003 | 0.00171 | 0.00040 | 0.00028 | 0.00002 |
| rs4242382 | 1 | rs7841060 | 2 | 0.00177 | 0.00002 | 0.00001 | 0.00407 | 0.00097 | 0.00004 | 0.00862 | 0.00165 | 0.00263 | 0.00448 |
| rs6983267 | 3 | rs7841060 | 2 | 0.01120 | 0.01027 | 0.00608 | 0.00027 | 0.00029 | 0.00000 | 0.02003 | 0.00521 | 0.00011 | 0.00319 |

**Supplementary Table 3**. Multiple regression analysis of primary signals from regions 1 - 4.[1]

| Region[1] | Variable | het OR | hom OR | p-value |
|---|---|---|---|---|
| 2 | rs7841060:G | 1.19 | 1.43 | 1.38E-12 |
| 4 | rs620861:C | 1.13 | 1.29 | 1.13E-08 |
| 3 | rs6983267:G | 1.20 | 1.44 | 3.14E-18 |
| 1 | rs4242382:A | 1.35 | 1.82 | 7.49E-21 |

[1] In order on chr8 and as described by Witte, J.S. Multiple prostate cancer risk variants on 8q24. Nat Genet 39, 579-80 (2007). Model included all 4 SNPs and covariates used in the single-SNP analyses.

**Supplementary Table 4.** Association results for all 8q24 SNPs from the present study

| Locus | Location | MAF[1] | p-value[2] |
|---|---|---|---|
| rs6999589 | 128154828 | 0.252\|0.251 | 5.73E-01 |
| rs1902431 | 128156258 | 0.236\|0.245 | 1.04E-01 |
| rs6470494 | 128157086 | 0.281\|0.309 | 4.49E-08 |
| rs1016342 | 128161637 | 0.471\|0.477 | 2.71E-01 |
| rs4871008 | 128162723 | 0.431\|0.399 | 6.16E-10 |
| rs7841060 | 128165659 | 0.211\|0.246 | 1.48E-13 |
| rs11993508 | 128169258 | 0.124\|0.117 | 1.75E-01 |
| rs1456316 | 128170030 | 0.356\|0.382 | 8.19E-07 |
| rs9656814 | 128170159 | 0.345\|0.344 | 9.57E-01 |
| rs17832285 | 128178175 | 0.175\|0.160 | 9.73E-04 |
| rs7825340 | 128178311 | 0.201\|0.215 | 1.25E-02 |
| rs7826337 | 128178756 | 0.305\|0.302 | 7.20E-01 |
| rs17765137 | 128179996 | 0.055\|0.044 | 5.22E-06 |
| rs7006409 | 128180611 | 0.323\|0.327 | 9.09E-01 |
| rs1378897 | 128191841 | 0.062\|0.055 | 4.30E-03 |
| rs1456305 | 128196434 | 0.127\|0.117 | 3.65E-03 |
| rs17446916 | 128232156 | 0.433\|0.442 | 8.97E-02 |
| rs7002343 | 128241210 | 0.351\|0.343 | 3.82E-01 |
| rs2124600 | 128241868 | 0.351\|0.354 | 9.21E-01 |
| rs2456461 | 128251633 | 0.409\|0.412 | 9.95E-01 |
| rs2466024 | 128257201 | 0.400\|0.403 | 9.81E-01 |
| rs2456449 | 128262163 | 0.343\|0.354 | 1.62E-01 |
| rs2456452 | 128266262 | 0.340\|0.346 | 4.63E-01 |
| rs2445610 | 128266270 | 0.358\|0.368 | 3.08E-01 |
| rs16902008 | 128266477 | 0.051\|0.042 | 5.60E-05 |
| rs2466031 | 128278791 | 0.406\|0.417 | 7.22E-02 |
| rs2466032 | 128279002 | 0.297\|0.308 | 8.03E-02 |
| rs2466035 | 128280411 | 0.313\|0.324 | 6.49E-03 |
| rs2445614 | 128281776 | 0.244\|0.247 | 1.43E-01 |
| rs9643217 | 128282415 | 0.078\|0.081 | 5.82E-01 |
| rs11991241 | 128288484 | 0.309\|0.327 | 1.11E-02 |
| rs7816475 | 128294622 | 0.231\|0.241 | 1.81E-01 |
| rs10087719 | 128298045 | 0.270\|0.275 | 7.96E-01 |
| rs10505481 | 128309583 | 0.237\|0.237 | 1.67E-01 |

| | | | |
|---|---|---|---|
| rs2044869 | 128316227 | 0.280\|0.275 | 2.15E-01 |
| rs17377068 | 128321036 | 0.211\|0.212 | 1.96E-01 |
| rs283741 | 128322175 | 0.437\|0.438 | 7.36E-01 |
| rs10104427 | 128325434 | 0.204\|0.200 | 5.78E-01 |
| rs6992922 | 128331109 | 0.212\|0.208 | 6.43E-01 |
| rs10099034 | 128336520 | 0.236\|0.221 | 1.28E-02 |
| rs4871014 | 128340162 | 0.251\|0.262 | 2.57E-02 |
| rs716889 | 128345535 | 0.034\|0.033 | 9.14E-01 |
| rs2122835 | 128347495 | 0.299\|0.310 | 3.65E-02 |
| rs11777807 | 128349990 | 0.429\|0.435 | 1.69E-01 |
| rs1011387 | 128351093 | 0.049\|0.047 | 6.34E-01 |
| rs283709 | 128358094 | 0.107\|0.107 | 8.76E-01 |
| rs283710 | 128358773 | 0.266\|0.280 | 4.15E-03 |
| rs4871780 | 128360760 | 0.425\|0.433 | 1.47E-01 |
| rs185852 | 128362648 | 0.209\|0.223 | 2.86E-03 |
| rs412835 | 128364892 | 0.416\|0.413 | 9.63E-01 |
| rs6984900 | 128373451 | 0.181\|0.164 | 2.18E-04 |
| rs283718 | 128376264 | 0.419\|0.422 | 5.61E-01 |
| rs17450934 | 128378019 | 0.121\|0.139 | 1.29E-06 |
| rs283720 | 128379147 | 0.287\|0.295 | 1.45E-01 |
| rs283721 | 128379675 | 0.134\|0.134 | 9.00E-01 |
| rs2007197 | 128380741 | 0.149\|0.133 | 2.36E-04 |
| rs6984136 | 128389320 | 0.044\|0.049 | 2.16E-02 |
| rs445114 | 128392363 | 0.370\|0.337 | 1.54E-09 |
| rs620861 | 128404855 | 0.372\|0.338 | 1.27E-10 |
| rs377649 | 128406423 | 0.491\|0.506 | 2.09E-02 |
| rs424281 | 128408608 | 0.452\|0.462 | 3.35E-01 |
| rs16902104 | 128410090 | 0.146\|0.165 | 1.20E-06 |
| rs587948 | 128410862 | 0.400\|0.371 | 4.44E-07 |
| rs687279 | 128413806 | 0.296\|0.284 | 7.21E-02 |
| rs672888 | 128414645 | 0.397\|0.379 | 3.32E-03 |
| rs10098985 | 128424201 | 0.410\|0.425 | 1.21E-02 |
| rs13281615 | 128424800 | 0.412\|0.396 | 1.02E-02 |
| rs16902124 | 128426400 | 0.040\|0.045 | 7.18E-03 |
| rs13267780 | 128426999 | 0.248\|0.225 | 2.01E-06 |
| rs11782735 | 128435786 | 0.144\|0.147 | 6.45E-01 |
| rs9693995 | 128437695 | 0.449\|0.431 | 4.93E-03 |
| rs4143118 | 128446650 | 0.439\|0.422 | 7.72E-03 |
| rs2060775 | 128447808 | 0.230\|0.212 | 3.52E-04 |
| rs16902126 | 128451539 | 0.421\|0.436 | 3.47E-02 |
| rs11776260 | 128451670 | 0.146\|0.149 | 6.24E-01 |
| rs1562430 | 128457034 | 0.415\|0.430 | 2.16E-02 |
| rs731900 | 128459842 | 0.182\|0.177 | 6.01E-01 |
| rs6986543 | 128465498 | 0.352\|0.360 | 1.42E-01 |
| rs896324 | 128465694 | 0.087\|0.094 | 1.29E-03 |
| rs7820981 | 128469358 | 0.420\|0.403 | 1.43E-02 |
| rs1562871 | 128470954 | 0.189\|0.192 | 3.56E-01 |
| rs7844673 | 128472696 | 0.066\|0.069 | 7.36E-01 |
| rs13258742 | 128617860 | 0.155\|0.156 | 7.85E-01 |
| rs4407842 | 128619305 | 0.444\|0.444 | 9.77E-01 |

| | | | |
|---|---|---|---|
| rs6470530 | 128628172 | 0.432\|0.438 | 4.02E-01 |
| rs10098729 | 128633980 | 0.095\|0.092 | 4.49E-01 |
| rs4620244 | 128635324 | 0.149\|0.153 | 4.08E-01 |
| rs11783049 | 128636470 | 0.099\|0.099 | 4.98E-01 |
| rs6470532 | 128637541 | 0.454\|0.445 | 2.02E-01 |
| rs7386167 | 128637894 | 0.351\|0.360 | 2.99E-01 |
| rs4733655 | 128638038 | 0.294\|0.296 | 4.02E-01 |
| rs4313118 | 128640941 | 0.185\|0.178 | 1.82E-01 |
| rs10097522 | 128642092 | 0.168\|0.162 | 5.81E-01 |
| rs12543106 | 128642480 | 0.299\|0.301 | 1.89E-01 |
| rs4733739 | 128647197 | 0.203\|0.200 | 9.22E-01 |
| rs4733766 | 128650508 | 0.327\|0.318 | 1.33E-01 |
| rs7824074 | 128658004 | 0.294\|0.300 | 4.79E-01 |
| rs7815738 | 128660144 | 0.143\|0.142 | 9.70E-03 |
| rs6470541 | 128667482 | 0.442\|0.446 | 7.49E-01 |
| rs10091329 | 128670554 | 0.102\|0.109 | 1.52E-01 |
| rs4129666 | 128672591 | 0.104\|0.099 | 1.91E-01 |
| rs7818319 | 128678824 | 0.096\|0.103 | 1.44E-01 |
| rs4451272 | 128682912 | 0.268\|0.271 | 5.98E-01 |
| rs4733597 | 128686794 | 0.376\|0.373 | 8.67E-01 |
| rs7005795 | 128692821 | 0.448\|0.447 | 8.83E-01 |
| rs7388104 | 128693606 | 0.472\|0.465 | 3.88E-01 |
| rs4551310 | 128696142 | 0.417\|0.430 | 6.21E-02 |
| rs6470552 | 128697783 | 0.075\|0.082 | 5.15E-02 |
| rs4314620 | 128699771 | 0.169\|0.167 | 7.23E-01 |
| rs13268507 | 128706649 | 0.442\|0.444 | 9.16E-01 |
| rs13271223 | 128706798 | 0.494\|0.492 | 9.40E-01 |
| rs6989963 | 128712413 | 0.427\|0.433 | 3.66E-01 |
| rs4130120 | 128714137 | 0.088\|0.089 | 3.05E-01 |
| rs10956381 | 128723297 | 0.411\|0.419 | 2.97E-01 |
| rs4733879 | 128724583 | 0.477\|0.475 | 1.57E-01 |
| rs11995971 | 128726459 | 0.493\|0.490 | 5.85E-01 |
| rs7817632 | 128729278 | 0.382\|0.377 | 3.12E-01 |
| rs4733616 | 128731277 | 0.184\|0.178 | 4.92E-01 |
| rs6985681 | 128732901 | 0.364\|0.366 | 7.53E-01 |
| rs4562278 | 128733772 | 0.198\|0.205 | 1.05E-01 |
| rs4385433 | 128734662 | 0.369\|0.374 | 6.16E-01 |
| rs7841193 | 128746713 | 0.491\|0.491 | 5.59E-01 |
| rs4593503 | 128748168 | 0.125\|0.123 | 7.69E-01 |
| rs12543549 | 128750181 | 0.275\|0.274 | 3.26E-01 |
| rs4733658 | 128759811 | 0.245\|0.248 | 7.02E-01 |
| rs11774100 | 128762703 | 0.349\|0.347 | 7.87E-01 |
| rs10505504 | 128769526 | 0.359\|0.358 | 7.70E-01 |
| rs7845292 | 128771858 | 0.426\|0.426 | 6.26E-01 |
| rs7840975 | 128774343 | 0.067\|0.068 | 8.24E-01 |
| rs4733676 | 128775901 | 0.425\|0.423 | 6.60E-01 |
| rs6470563 | 128777752 | 0.107\|0.112 | 5.66E-01 |
| rs4733677 | 128781003 | 0.233\|0.233 | 9.99E-01 |
| rs12547643 | 128782355 | 0.350\|0.346 | 7.67E-01 |
| rs16902328 | 128783055 | 0.061\|0.065 | 1.08E-02 |

| | | | |
|---|---|---|---|
| rs9642880 | 128787250 | 0.465\|0.466 | 9.93E-01 |
| rs10505505 | 128808953 | 0.054\|0.057 | 6.00E-01 |
| rs4645943 | 128816653 | 0.047\|0.047 | 9.14E-01 |

[1] controls|cases
[2] 2df genotype score test

## Supplementary Note

## *Studies*

### 1. CGEMS Stage 1 -- Prostate, Lung, Colon and Ovarian (PLCO)

For PLCO study description, please refer to the Supplementary Methods from Yeager et al. 2007[2].

### 2. CGEMS Stage 2 – Follow-up #1

For Health Professionals Follow-up Study (HPFS), Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study (ATBC), American Cancer Society Cancer Prevention Study II Nutrition Cohort (CPS-II), and CeRePP French Prostate Case-Control Study (CeRePP), please refer to the Supplementary Methods Thomas et al. 2008[3].

### 3. Additional Studies included within Stage 3:

#### a. The Cancer Prostate in Sweden study (CAPS)

The Cancer Prostate in Sweden study (CAPS) is a large population-based Swedish case-control study[4]. Prostate cancer patients were identified and recruited from four of the six regional cancer registries in Sweden. Compulsory reporting of all malignant diseases diagnosed by both clinicians and pathologists ensures an essential complete cases ascertainment. Eligible case subjects were all men under 80 years of age with pathological or cytological verified adenocarcinoma of the prostate (ICD-10: C61), diagnosed between July, 2001 and October, 2003. Among 3,648 identified prostate cancer case subjects, 3,161 (87%) agreed to participate. DNA samples from blood and TNM stage, Gleason grade (biopsy), and PSA levels at diagnosis were available for 2,893 patients (91%). These case subjects were classified as having advanced disease if they met any of the following criteria: T3/4, N+, M+, Gleason score sum ≥ 8, or PSA > 50 ng/ml; otherwise, they were classified as localized.

Control subjects were recruited concurrently with case subjects. They were randomly selected from the Swedish Population Registry, and frequency matched according to the expected age distribution of cases (groups of five-year interval) and geographical region. Control subjects with a history of prostate cancer were excluded from the study. A total of 3,153 controls were invited and 2,149 (68%) agreed to participate. DNA samples from blood were available for 1,781 control subjects (83%). Serum PSA level was measured for all control subjects but was not used as exclusion criteria.

**b. The Multiethnic Cohort Study (MEC)**

The Multiethnic Cohort Study (MEC) is a population-based prospective cohort study that was initiated between 1993 and 1996 and includes subjects mainly from five self-reported populations – African-Americans and Latinos primarily from California (mainly Los Angeles) and Native Hawaiians, Japanese-Americans, and European Americans primarily from Hawaii[5]. State driver's license files were the primary sources used to identify study subjects in Hawaii and California. Additionally, in Hawaii, state voter's registration files were used, and, in California, Health Care Financing Administration (HCFA) files were used to identify additional African American men.

All participants (n=215,251) returned a 26-page self-administered baseline questionnaire that obtained general demographic, medical and risk factor information. In the cohort, incident cancer cases are identified annually through cohort linkage to population-based cancer Surveillance, Epidemiology, and End Results (SEER) registries in Hawaii and Los Angeles County as well as to the California State cancer registry. Information on stage and grade of disease are also obtained through the SEER registries.

Blood sample collection in the MEC began in 1994 and targeted incident prostate cancer cases and a random sample of study participants to serve as controls for genetic analyses. This nested prostate cancer case-control study in the MEC consists of 3,000 invasive prostate cancer cases and 3,000 age-matched controls (in 5-year age groups), including 750 cases and 750 controls of ancestry, with the majority of predominantly European-descent. This study was approved by the Institutional Review Boards at the University of Southern California and at the University of Hawaii and informed consent was obtained from all study participants.

**c. The European Prospective Investigation into Cancer and Nutrition (EPIC)**

The European Prospective Investigation into Cancer and Nutrition (EPIC) has been established to investigate the relationship between diet, nutritional status, lifestyle and environmental factors and the incidence of cancer and other chronic diseases. It is a large ongoing prospective study that recruited 520,000 subjects, 370,000 women and 150,000 men, between years 1992-2000, in 23 regional centers in ten European countries: Denmark, France, Italy, Germany, The Netherlands, Norway, Spain, Sweden and the United Kingdom.

The study subjects' recruitment initiated in 1992. Detailed information on diet and lifestyle was collected using standardized questionnaires, anthropometric characteristics; blood pressure and pulse rate were measured at recruitment. Blood samples were also collected from approximately 400,000 of these subjects

(37% males); plasma, serum, erythrocytes and leucocytes were aliquoted and stored in liquid nitrogen in bio repositories. Detailed EPIC recruitment procedure and collection of samples is described elsewhere[6].

EPIC contributed 674 prostate cases and 1004 controls towards the CGEMS study. Cases and controls were matched based on study center, length of follow-up and age at blood collection (+/- 6 months), fasting status and time of the day of blood drawing (+/- 1 hour).

**c. Johns Hopkins University (JHU)**

The Johns Hopkins University (JHU) study population was described in detail elsewhere[7]. Briefly, the prostate cancer patients were men of European descent (by self report) who underwent radical prostatectomy for treatment of prostate cancer at The Johns Hopkins Hospital from January 1, 1999, through December 31, 2006. Normal seminal vesicle tissue that was obtained and frozen at the time of surgery was used to isolate DNA for genotyping of case patients. During the same time period, men undergoing screening for prostate cancer at The Johns Hopkins Hospital and The Johns Hopkins University Applied Physics Lab (Columbia, MD) were asked to participate as control subjects. Serum prostate-specific antigen (PSA) levels, digital rectal examination (DRE) results, and demographic information were available for these subjects. Men of European descent (by self report) met our inclusion criteria as control subjects for this study: normal DRE, PSA levels less than 4.0 ng/mL, and age older than 55 years.

**d. CONOR**

CONOR[8] is a collaboration of six population-based cohorts in Norway, including approximately 180,000 participants of both sexes where information on the participants is collected with regular intervals. Core variables have been established through questionnaires, clinical measurements and venous blood sampling. Among the cohorts, extracted DNA for this study was available from two component cohorts (the HUNT and Tromsø studies), including approximately 45,000 men who are being followed up for prostate cancer incidence. Incident prostate cancer cases were ascertained by linkage to the Norwegian Cancer Registry, and controls were matched to the cases by age and study cohort. In this study, 606 cases and 622 controls were included in this analysis.

## Supplementary Methods

Cancer Genetic Markers of Susceptibility (CGEMS) Prostate Cancer Follow-up Scan 2

The CGEMS cancer prostate whole-genome scan began by genotyping approximately 550,000 SNPs from the Illumina HumanHap550 assay in more than 1,100 prostate cancer patients and an equivalent number of controls from the PLCO Cancer Screening Trial. The materials and methods of the initial genome-wide scan have been reported[2] and are also available at http://cgems.cancer.gov/data/. Based on the analysis of the initial scan, a follow-up scan included 29,018 of the most promising SNPs to type in an additional 4,020 prostate cancer cases and 4,028 controls drawn from four additional studies using a custom Illumina iSelect™ assay chip[3]. A second follow up, reported here, was performed with the primary purpose of fine mapping the most promising association results from the previous stages assaying 6,612 SNPs using a custom Illumina iSelect™ assay chip in 9,135 prostate cancer cases and 10,286 controls. The component study design, SNP selection, assay performance, and analysis methods for this second follow-up scan are described below.

### *Follow-up Design*

A total of 7,034 SNPs were chosen for second round follow-up genotyping and were chosen based on several criteria (Table A).

A. Prostate follow-up #1 replication: The majority (51%) of SNPs were chosen for testing loci that were observed to be noteworthy in prostate follow-up #1 (reported elsewhere). For each of 27,157 SNPs genotyped in prostate follow-up #1[3], a categorization was determined as follows:

   - Region Type 0 were designated such that:
     SNPs denoted as being in this type of region were observed to have a $-\log_{10}$ p value less than 3 from the trichotomous analysis. These SNPs were *not* taken into follow-up #2; the one exception was a SNP (rs4857841) that was included in the follow-up due to having a $-\log_{10}$ p value from the dichotomous analysis less than 4, although the dichotomous p value did not meet the threshold.

   - Region Type 1 were designated such that:
     The observed $-\log_{10}$ p value from the trichotomous analysis was greater than 3 and the $-\log_{10}$ p value from the dichotomous analysis was less than $10*(-\log_{10}$ p value from the trichotomous analysis - 3.525). Region Type 1 also contains the exception mentioned in the description of region type 0 (rs4857841).

- Region Type 2 were designated such that:
  The observed $-\log_{10}$ p value from the dichotomous analysis was less than $5*(-\log_{10}$ p value from the trichotomous analysis - 4.5) and $-\log_{10}$ p value from the dichotomous analysis was greater than $10*(-\log_{10}$ p value from the trichotomous analysis - 3.525)

- Region Type 3 were designated such that:
  The observed $-\log_{10}$ p value from the dichotomous analysis was greater than $5*(-\log_{10}$ p value from the trichotomous analysis - 4.5)

For each region type (1 – 3), the following strategy was used to optimally explore each region:

- Region 1 – 58 regions. Initial region bounds were defined by using the 0.2cM HapMap recombination data for the most significant SNP within the region. Regions were then tagged at an $r^2$ of 0.6 using HapMap CEU, with all significant SNPs ($p < 10^{-3}$) serving as obligate-includes; final tags were chosen for follow-up #2 inclusion if they were observed to be correlated with an $r^2$ of $\geq 0.8$ in HapMap[9,10] CEU, YRI, JPT+CHB with the obligate-includes.

- Region 2 – 54 regions. Initial region bounds were defined by using the 0.2cM HapMap recombination data for the most significant SNP within the region. Regions were then tagged at an $r^2$ of 0.4 using HapMap CEU, with all significant SNPs ($p < 10^{-3}$) serving as obligate-includes; final tags were chosen for follow-up #2 inclusion if they were observed to be correlated with an $r^2$ of $\geq 0.8$ in HapMap[9,10] CEU, YRI, JPT+CHB with the obligate-includes.

- Region 3 – 24 regions. Initial region bounds were defined by using the 0.2cM HapMap recombination data for the most significant SNP within the region. Regions were then tagged at a D' of 0.6 using HapMap CEU, with all significant SNPs ($p < 10^{-3}$) serving as obligate-includes; final tags were chosen for follow-up #2 inclusion if they were observed to be correlated with an $r^2$ of $\geq 0.8$ in HapMap[9,10] CEU, YRI, JPT+CHB with the obligate-includes.

Tag SNP selection was performed using the GLU software package (http://code.google.com/p/glu-genetics/) using the HapMap CEU, JPT+CHB and YRI data.

B. Failed follow-up #1: From the prostate follow-up #1 Illumina iSelect, attempted assays for SNPs that were excluded from the design, failed *in silico* design, manufacturing or QC[3]. These include 369 SNPs located on Chromosome X that were incorrectly excluded from follow up #1 during the follow up design.

C. Population substructure: 1,408 SNPs were chosen to monitor population stratification[11].

D. 8q24 prostate regions: 313 SNPs were included to saturate a ~600kb region of chromosome 8q24 (chr8: 128,154,828 – 128816653) for numerous criteria, including:
   a. Resequencing[12] of regions 1[3,13] and 3[12]
   b. SNPs included in prostate follow-up #1[3]
   c. Fine mapping based on tag SNPs of several regions of interest

E. Candidate genes/regions/SNPs: SNPs added based on candidate gene hypotheses or alternative analyses.

**Table A. Follow-up #2 SNPs**

| Hypothesis | Ordered | Manufactured | Passed QC |
|---|---|---|---|
| Follow-up #1 replication | | | |
| *Region Type 1* | *120* | *111* | *110* |
| *Region Type 2* | *554* | *519* | *514* |
| *Region Type 3* | *2,920* | *2,777* | *2,752* |
| Failed follow-up #1 | 1,695 | 1,591 | 1,588 |
| Population Stratification | 1,474 | 1,400 | 1,399 |
| 8q24 | 313 | 297 | 293 |
| Other SNPs | *27* | *23* | *23* |
| **Total*** | **7,033** | **6,652** | **6,613** |

* Some SNPs were included based on several criteria. Thus the total number reflects the number of unique SNPs and will generally be less than the sum of each of the individual hypotheses.

**This paper reports *only* on the SNPs tested in section above D (b. and c.) and explores Regions 2, the breast cancer region and bladder cancer regions (see text for description).**

## *Genotype Quality control*

### Assessment of Call Rates

A total of 6,652 SNP genotype assays were attempted on the 22,081 DNA samples including 22,057 study samples and 24 CEPH samples using the Illumina iSelect. Samples that did not meet a 90% completion threshold were excluded from further analysis. See Table B for the number of samples from each cohort that was excluded based on these criteria. The remaining 20,707 DNA samples were retained for the subsequent analyses.

SNPs were excluded based on the following criteria: 1) The assay failed manufacturing at Illumina or 2) The assay exhibited low completion rate in the laboratory. A total of 38 SNPs failed to provide reliable genotype results due to either no call or low call rates (<90%; see Table A for distribution among SNP categories of inclusion).

**Table B. Samples genotyped**

| Study | Attempted | Failed QC | Passed QC |
|---|---|---|---|
| PLCO | 2,256 | 198 | 2,058 |
| HPFS | 1,296 | 30 | 1,266 |
| ATBC | 2,004 | 150 | 1,854 |
| CPS-II (blood) | 2,640 | 69 | 2,571 |
| CPS-II (buccal) | 1,104 | 156 | 948 |
| CeRePP | 2,160 | 96 | 2,064 |
| CAPS | 4,200 | 459 | 3,741 |
| MEC | 1,536 | 44 | 1,492 |
| EPIC | 1,726 | 16 | 1,710 |
| JHU | 1,779 | 102 | 1,677 |
| CONOR | 1,356 | 30 | 1,326 |
| **Total** | 22,057 | 1,350 | 20,707 |

**Table C. Subject counts by study that passed sample-level genotype QC**

| | | | Cases | | | |
|---|---|---|---|---|---|---|
| Study | QC/ Unknown phenotype | Controls | All | Non-aggressive tumor | Aggressive tumor | Unknown stage tumor |
| PLCO | 38 | 931 | 975 | 425 | 550 | 0 |
| HPFS | 0 | 598 | 611 | 427 | 115 | 69 |
| ATBC | 0 | 870 | 907 | 527 | 245 | 135 |
| CPS-II (blood) | 64 | 1,208 | 1,212 | 654 | 427 | 131 |
| CPS-II (buccal) | 0 | 448 | 448 | 7 | 410 | 31 |
| CeRePP | 0 | 970 | 1,016 | 238 | 778 | 0 |
| CAPS | 0 | 1,375 | 2,226 | 999 | 1,190 | 37 |
| MEC | 0 | 731 | 725 | 365 | 239 | 121 |
| EPIC | 0 | 995 | 685 | 462 | 223 | 0 |
| JHU | 1 | 600 | 1,002 | 495 | 504 | 3 |
| CONOR | 0 | 670 | 609 | 0 | 0 | 609 |
| **Total** | 103 | 9,396 | 10,416 | 4,599 | 4,681 | 1,136 |

**Assessment of unique subjects**

After removal of sample and locus data due to low completion rates, genotypes for each sample that appeared in duplicate were merged to form consensus genotypes for each study subject; for any observed genotype discordances were henceforth considered as missing observations. Table C contains the detailed numbers for each study of the distribution of subjects by phenotype.

**Analysis of duplicate DNA samples**

The genotype concordance/reproducibility rate for SNP assays was evaluated using the 2,976 pairs of known duplicated DNA samples. These pairs of samples were separate aliquots from the same DNA preparation and all met quality control criteria requested for the other samples, thereby, providing reliable data for comparison. An average discordance rate of 0.03% was observed. No SNPs or samples were excluded from further analysis as a result of this analysis of *known* duplicates. Table D shows the individual study discordance rates in more detail.

**Table D. Intra-cohort sample genotype concordance**

| | | | | | | | Discordance Rate | |
|---|---|---|---|---|---|---|---|---|
| **Study** | **Pairs** | **Assays** | **Subjects** | **Discordant Genotypes** | **Concordant Genotypes** | **Total Comparisons** | **Mean** | **Max** |
| CPS-II (blood) | 355 | 109 | 35 | 2,166 | 2,237,370 | 2,239,536 | 0.100% | 2.50% |
| CPS-II (buccal) | 83 | 84 | 36 | 8 | 506,081 | 506,089 | 0.002% | 0.02% |
| ATBC | 251 | 114 | 43 | 830 | 1,575,327 | 1,576,157 | 0.050% | 1.00% |
| CeRePP | 315 | 113 | 39 | 861 | 1,971,908 | 1,972,769 | 0.040% | 1.00% |
| HPFS | 134 | 87 | 34 | 74 | 859,437 | 859,511 | 0.009% | 0.80% |
| CONOR | 137 | 61 | 20 | 36 | 886,080 | 886,116 | 0.004% | 0.50% |
| MEC | 33 | 66 | 33 | 51 | 208,678 | 208,729 | 0.020% | 0.40% |
| JHU | 249 | 112 | 42 | 492 | 1,556,024 | 1,556,516 | 0.030% | 0.40% |
| EPIC | 24 | 48 | 27 | 33 | 133,611 | 133,644 | 0.020% | 0.50% |
| CAPS | 1,075 | 201 | 65 | 1,531 | 6,535,579 | 6,537,110 | 0.020% | 0.90% |
| PLCO | 320 | 123 | 45 | 348 | 2,011,163 | 2,011,511 | 0.020% | 0.60% |
| **Total** | 2,976 | 1,118 | 419 | 6,430 | 18,481,258 | 18,487,688 | 0.030% | 2.50% |

Concordance analysis among all possible pairs of subjects also revealed unexpected pairs with nearly identical genotypes. An individual assayed multiple times (N) will generate N(N-1)/2 pairs. For example, in CONOR there are 3 CEPH subject repeated 5 times, 1 CEPH sample repeated 14 times, and 16 study samples in duplicate. This results in 137 duplicate pairs, but comprises only 61 assays and 20 individuals.

Within-study pairs are likely to reflect sample-handling errors, while inter-study duplicates are possibly monozygotic twins or, more likely, individuals enrolled in more than one study. 27 pairs of such subjects were found and were verified to have nearly identical age and phenotypes. Table E shows the discordance rates per-study for unexpected duplicate pairs. Although a vast minority if the total data set, one or both subjects from each pair was excluded to prevent repeated measures within the analysis.

**Table E. Inter-cohort sample genotype concordance**

| | | | Genotype Counts | | | Discordance Rate | |
|---|---|---|---|---|---|---|---|
| Study1 | Study2 | Pairs | Discordant Genotypes | Concordant Genotypes | Total Comparisons | Mean | Max |
| CPSII-BLOOD | HPFS | 5 | 9 | 31466 | 31475 | 0.029% | 0.048% |
| CPSII-BUCCAL | HPFS | 1 | 1 | 6065 | 6066 | 0.016% | 0.016% |
| EPIC | EPIC | 3 | 2 | 19094 | 19096 | 0.010% | 0.031% |
| PLCO | CPSII-BLOOD | 5 | 64 | 30551 | 30615 | 0.217% | 1.053% |
| PLCO | CPSII-BUCCAL | 3 | 0 | 18162 | 18162 | 0.000% | 0.000% |
| PLCO | HPFS | 3 | 3 | 18737 | 18740 | 0.016% | 0.032% |
| PLCO | JHU | 1 | 0 | 6205 | 6205 | 0.000% | 0.000% |
| CAPS | EPIC-Sweden | 6 | 9 | 36298 | 36307 | 0.025% | 0.033% |
| **Total** | | 27 | 88 | 166578 | 166666 | 0.053% | 1.053% |

**Hardy –Weinberg Proportions in control DNA**

Genotype data were tested for deviation from Hardy-Weinberg proportions using an exact test[14]. The analysis was conducted in each cohort's control group. Significant deviations were observed for an average of 5.44% for control group and 5.77% for case group of all SNPs at the level of $p<0.05$, and 0.62% for control group and 0.61% for case group at $p<0.001$. Table F contains the proportion of SNPs per study that deviate from Hardy-Weinberg proportions. None of these SNPs were excluded from analysis since significant departures from the expected proportions are not unexpected when fine-mapping previously associated regions. Genotype tests for association applied to such data are valid in the presence of departure from Hardy-Weinberg proportions, although with potentially reduced power when these deviations are due to systematic genotyping errors with comparable effects among cases and controls.

**Table F. SNPs exhibiting deviation from Hardy-Weinberg Proportions**

| | Cases | | Controls | |
|---|---|---|---|---|
| **Study** | **P<0.001** | **P<0.05** | **P<0.001** | **P<0.05** |
| ATBC | 0.7888% | 5.6368% | 0.6245% | 4.6343% |
| CONOR | 0.5657% | 5.5762% | 0.6304% | 5.1398% |
| CPSII-BLOOD | 0.9241% | 5.7096% | 1.0561% | 5.3630% |
| CPSII-BUCCAL | 0.3319% | 4.9286% | 0.4481% | 5.5758% |
| EPIC | 0.3897% | 5.6827% | 0.5520% | 5.6178% |
| CeRePP | 0.5578% | 5.5291% | 0.5578% | 5.2338% |
| HPFS | 0.5210% | 4.5751% | 0.6024% | 4.5751% |
| JHU | 0.4884% | 5.8613% | 0.3419% | 4.6076% |
| MEC | 1.0374% | 6.7841% | 1.0868% | 7.5086% |
| PLCO | 0.5230% | 7.2234% | 0.4576% | 6.7331% |
| CAPS | 0.5967% | 6.0501% | 0.4973% | 4.8898% |

**Subject exclusions**

Subjects with valid genotypes were excluded from analysis based on the following (summarized in Table G):

1. For pairs of subjects with unanticipated duplicate genotype data, the decision as to which subject to retain, if any, was based on the following criteria:
    a. If both subjects were found in the same study, both were excluded unless other evidence could be found to convincingly determine the identity of one (such as another known duplicate)
    b. Any subject from an earlier phase of the scan was retained over that of a new subject in follow-up #2; and
    c. If two subjects were first seen in follow #2, the one with higher genotyping completion was retained for analysis.
2. Self-reported race, where only subjects that self-reported as being of European ancestry were retained, regardless of their imputed race.
3. Imputed race, where only subject with imputed European background were retained, regardless of their self-reported race.
4. Sparse groups. Only 2 PLCO subjects were available from one study center and were excluded. Also, only 5 non-aggressive cases were part of the CPS-II buccal group, so they too were also excluded.
5. Missing covariates. These subjects were missing one or more covariates necessary for association analysis.

**Table G. Excluded Subjects**

| Study | Inter-study Duplicates | Non-European Origin* | Sparse Group | Missing Covariates | PI Exclude | Total |
|---|---|---|---|---|---|---|
| PLCO | 0 | 5 | 2 | 36 | 1 | 44 |
| CPS-II | 8 | 28 | 0 | 31 | 33 | 100 |
| ATBC | 0 | 3 | 0 | 0 | 0 | 3 |
| CeRePP | 0 | 36 | 0 | 0 | 0 | 36 |
| HPFS | 9 | 16 | 0 | 0 | 0 | 25 |
| EPIC | 0 | 6 | 0 | 0 | 0 | 6 |
| MEC | 0 | 87 | 0 | 0 | 0 | 87 |
| CONOR | 0 | 11 | 0 | 0 | 0 | 11 |
| JHU | 1 | 160 | 0 | 1 | 0 | 162 |
| CAPS | 0 | 20 | 0 | 0 | 0 | 20 |
| **Total** | 18 | 372 | 2 | 68 | 34 | 494 |

* Non-European origin is defined as having less than 0.80 estimated European admixture as estimated using the STRUCTURE program using with the HapMap Phase II populations.

After the exclusion of subjects based on various criteria, the total number of cases and controls for association analyses were 10,286 and 9,135, respectively. Table H contains the final subject counts per study for each phenotypic state.

**Table H. Final subject counts for association analysis**

| | | Cases | | | |
|---|---|---|---|---|---|
| **Study** | **Controls** | **All Cases** | **Non-aggressive tumor** | **Aggressive tumor** | **Unknown stage tumor** |
| PLCO | 927 | 973 | 425 | 548 | 0 |
| CPS-II | 1,644 | 1,636 | 651 | 826 | 159 |
| ATBC | 868 | 906 | 527 | 244 | 135 |
| CeRePP | 952 | 998 | 235 | 763 | 0 |
| HPFS | 589 | 595 | 417 | 109 | 69 |
| EPIC | 991 | 683 | 462 | 221 | 0 |
| CONOR | 662 | 606 | 0 | 0 | 0 |
| MEC | 687 | 682 | 341 | 226 | 115 |
| JHU | 451 | 990 | 489 | 498 | 3 |
| CAPS | 1,364 | 2,217 | 1,184 | 996 | 37 |
| **Total** | 9,135 | 10,286 | 4,731 | 4,431 | 518 |

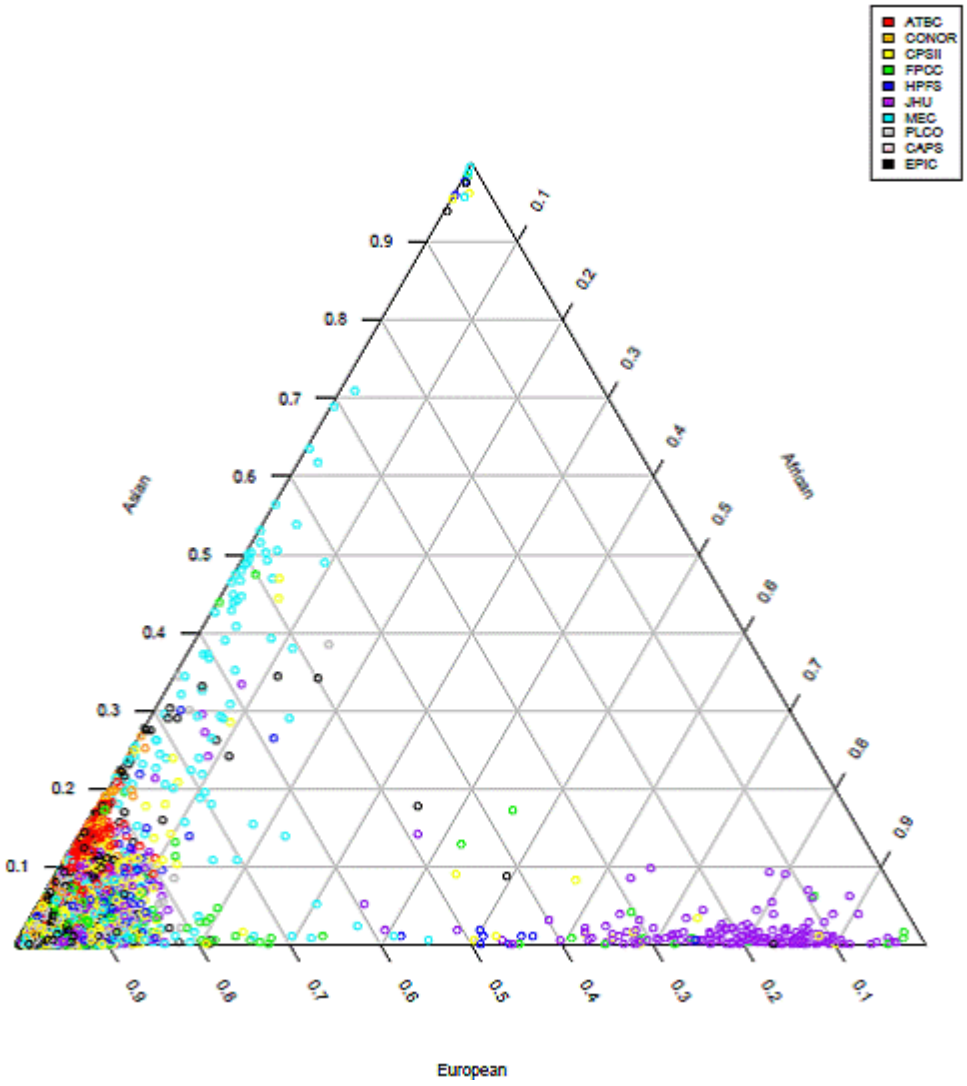
**Admixture estimation**

1,400 SNPs were used for the detection of population structure. In an attempt to maximize genetic homogeneity, subjects with significant estimated non-European ancestry were excluded from analysis. This estimation was done using the STRUCTURE program by merging the genotypes from the follow-up studies with those of the reference HapMap populations. The number of clusters (the “k” parameter) was set to three and the CEU, YRI and JPT+CHB samples were each specified to a different cluster schematically representing populations of European, African and Asian origin, respectively. The origin of the CGEMS samples was left unspecified. A total of 372 subjects (1.8%) were estimated to have less than 80% European ancestry and were excluded from analysis (Figure A). All individuals that had at greater than 80% European ancestry were retained for the replication study, regardless of their reported origin (Table I).

Table I. Self-reported and imputed race per study

| | | Imputed race | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Study** | **Self-Reported Race** | **CEU** | **Admixed CEU** | **Admixed YRI** | **Asian** | **Asian/ CEU** | **CEU/ YRI** | **YRI** |
| ATBC | Caucasian | 1,774 | 1 | | | 2 | | |
| CONOR | Caucasian | 1,268 | 5 | | | 6 | | |
| CPSII | African-American | 2 | | | | | 9 | 2 |
| | Asian | 1 | | | 2 | | | |
| | Caucasian | 3283 | 5 | | | 1 | 2 | |
| | Caucasian/Hispanic | | 1 | | | | | |
| | Hispanic | 2 | 1 | | | 4 | | |
| | Other | | | | | 1 | | |
| | None | 64 | | | | | | |
| EPIC | Caucasian | 1674 | | | 1 | 3 | 2 | |
| FPCC | Caucasian | 1950 | 8 | | 1 | 2 | 20 | 5 |
| HPFS | Caucasian | 1129 | 3 | | | 2 | | |
| | Other | 64 | | | 2 | 1 | 8 | |
| JHU | African-American | | | 10 | | | 65 | 73 |
| | Caucasian | 1443 | 2 | | | 5 | 4 | 1 |
| MEC | Caucasian | 1369 | 13 | | 3 | 63 | 8 | |
| PLCO | Caucasian | 1904 | 1 | | | 2 | 1 | |
| | Other | 35 | | | | 1 | | |
| CAPS | Caucasian | 3581 | 2 | | 1 | 14 | 2 | 1 |
| **Total** | | 19543 | 42 | 10 | 10 | 107 | 121 | 82 |

**Figure A. Admixture analysis using 1,400 SNPs[11]**




**Principal components analyses**

Principal components analyses were performed using the 1,400 SNPs included for population stratification. These results were based on the remaining subjects after removal of the admixed individuals as detected above.

A Wilcoxon rank test was performed to check correlations with the case/control status for the top 5 eigenvectors. The result is summarized in Table J below. PC1 in JHU and top 3 PCs in CAPS show significant differences between cases and controls. These are potential adjustors for association test.

Table J. Wilcoxon Rank test for association between phenotype and each of the top 5 principle components (PC1-5)

| Study | PC1 | PC2 | PC3 | PC4 | PC5 |
|---|---|---|---|---|---|
| ATBC | 6.90E-01 | 2.22E-01 | 4.93E-01 | 1.93E-01 | 6.56E-01 |
| CONOR | 3.42E-01 | 1.34E-01 | 4.97E-01 | 7.76E-01 | 2.87E-01 |
| CPSII | 6.81E-01 | 1.39E-01 | 7.01E-01 | 8.32E-01 | 3.50E-01 |
| CeRePP | 8.49E-01 | 9.95E-01 | 2.75E-01 | 8.98E-01 | 7.65E-01 |
| HPFS | 3.87E-01 | 7.66E-01 | 9.22E-01 | 9.58E-01 | 6.20E-01 |
| JHU | 7.92E-02 | 8.58E-01 | 4.82E-01 | 9.75E-01 | 5.70E-01 |
| MEC | 5.15E-01 | 9.50E-01 | 5.44E-01 | 8.30E-01 | 7.17E-01 |
| PLCO | 8.22E-01 | 1.32E-01 | 9.73E-01 | 8.09E-01 | 5.30E-01 |
| CAPS | 3.32E-05 | 4.15E-03 | 7.48E-02 | 3.06E-01 | 3.43E-01 |
| EPIC | 5.09E-01 | 6.02E-01 | 6.76E-01 | 1.86E-01 | 1.12E-01 |

## Association Analysis

Single-SNP logistic regression tests for association with comparing prostate cancer cases to controls were conducted using the GLU software package (http://code.google.com/p/glu-genetics/). Models was adjusted for study, study center (when applicable), and for significant principal components per study (see below). A polytomous, non-ordinal model for comparisons between non-aggressive and aggressive cases vs. controls was similarly performed. Genotype effects were modeled using both an unconstrained genotype effects (two parameters, one for heterozygote genotypes and another for non-reference homozygote genotypes), as well as a "trend" test which assumes a multiplicative relationship between the heterozygote and non-reference homozygote odds ratios. A score test was performed on all genetic parameters in each model to determine statistical significance, with two degrees of freedom for unconstrained genotype effects and one degree of freedom for trend effects.

## Genotype concordance

In order to reduce the likelihood that the association results that we obtained from regions 2 and 4 were due to genotyping artifacts, we ran TaqMan assays on a subset of individuals from the CPSII (n = 379) and PLCO (n = 621) studies for the 2 most highly-associated SNPs from each region. Table K shows the concordance rates per locus and study between TaqMan and iSelect genotypes.

**Table K. Concordance between TaqMan and iSelect genotypes**

| locus | region | **CPSII (n=379)** | **PLCO (n=601)** |
|---|---|---|---|
| rs4871008 | 2 | 0.9973 | 0.9916 |
| rs7841060 | 2 | 0.9973 | 0.9983 |
| rs445114 | 4 | 0.9918 | 0.9932 |
| rs620861 | 4 | 0.9947 | 0.9933 |
| | average | 0.9953 | 0.9941 |

**REFERENCES**


1. Witte, J.S. Multiple prostate cancer risk variants on 8q24. *Nat Genet* **39**, 579-80 (2007).
2. Yeager, M. et al. Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Nat Genet* **39**, 645-9 (2007).
3. Thomas, G. et al. Multiple loci identified in a genome-wide association study of prostate cancer. *Nat Genet* (2008).
4. Zheng, S.L. et al. Cumulative association of five genetic variants with prostate cancer. *N Engl J Med* **358**, 910-9 (2008).
5. Kolonel, L.N. et al. A multiethnic cohort in Hawaii and Los Angeles: baseline characteristics. *Am J Epidemiol* **151**, 346-57 (2000).
6. Riboli, E. et al. European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. *Public Health Nutr* **5**, 1113-24 (2002).
7. Zheng, S.L. et al. Association between two unlinked loci at 8q24 and prostate cancer risk among European Americans. *J Natl Cancer Inst* **99**, 1525-33 (2007).
8. Naess, O. et al. Cohort profile: cohort of Norway (CONOR). *Int J Epidemiol* **37**, 481-5 (2008).
9. Project, T.I.H. The International HapMap Project. *Nature* **426**, 789-96 (2003).
10. Consortium, T.I.H. A haplotype map of the human genome. *Nature* **437**, 1299-320 (2005).
11. Yu, K. et al. Population substructure and control selection in genome-wide association studies. *PLoS ONE* **3**, e2551 (2008).
12. Yeager, M. et al. Comprehensive resequence analysis of a 136 kb region of human chromosome 8q24 associated with prostate and colon cancers. *Hum Genet* **124**, 161-70 (2008).
13. Amundadottir, L.T. et al. A common variant associated with prostate cancer in European and African populations. *Nat Genet* **38**, 652-8 (2006).
14. Wigginton, J.E., Cutler, D.J. & Abecasis, G.R. A note on exact tests of Hardy-Weinberg equilibrium. *Am J Hum Genet* **76**, 887-93 (2005).