

AWARD NUMBER: W81XWH-12-1-0183

TITLE: Metabolomic Footprints of Lethal Versus Indolent Prostate Cancer

PRINCIPAL INVESTIGATOR: Jing Ma

CONTRACTING ORGANIZATION: Brigham and Women's Hospital  
Boston, MA 02115

REPORT DATE: October 2014

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command  
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;  
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

**REPORT DOCUMENTATION PAGE**Form Approved  
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

<b>1. REPORT DATE</b> October 2014		<b>2. REPORT TYPE</b> Annual		<b>3. DATES COVERED</b> 30 Sep 2013 - 29 Sep 2014	
<b>4. TITLE AND SUBTITLE</b> Metabolomic Footprints of Lethal Versus Indolent Prostate Cancer				<b>5a. CONTRACT NUMBER</b>	
				<b>5b. GRANT NUMBER</b> W81XWH-12-1-0183	
				<b>5c. PROGRAM ELEMENT NUMBER</b>	
<b>6. AUTHOR(S)</b> Jing Ma, Clary Clish, Jorge Chavarro, Weiliang Qiu, Yin Cao, Meng Yang, Changzheng Yuan  E-Mail: hpjma@channing.harvard.edu				<b>5d. PROJECT NUMBER</b>	
				<b>5e. TASK NUMBER</b>	
				<b>5f. WORK UNIT NUMBER</b>	
<b>7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)</b> Brigham and Women's Hospital 75 Francis Street, Boston, MA 02115, USA				<b>8. PERFORMING ORGANIZATION REPORT NUMBER</b>	
<b>9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)</b> U.S. Army Medical Research and Materiel Command Fort Detrick, Maryland 21702-5012				<b>10. SPONSOR/MONITOR'S ACRONYM(S)</b>	
				<b>11. SPONSOR/MONITOR'S REPORT NUMBER(S)</b>	
<b>12. DISTRIBUTION / AVAILABILITY STATEMENT</b> Approved for Public Release; Distribution Unlimited					
<b>13. SUPPLEMENTARY NOTES</b>					
<b>14. ABSTRACT</b> The current study proposed to use a targeted, LC-MS-based metabolite profiling platform to measure and compare metabolic profiles of prediagnostic blood samples collected from men subsequently diagnosed with PC and sample of men who remained cancer-free in Physicians' Health Study (PHS). And test whether these relationships are independent of the known metabolic risk factors (overweight/obese, insulin marker C-peptide, insulin-like growth factor I (IGF-I), IGF binding protein 3, (IGFBP-3), and adiponectin) as well as the clinical characteristics defined as the D'Amico risk. In the original protocol, we plan to measure samples of PC cases from both HPFS and PHS. However, the HPFS team has received separate grant for metabolomics measurement. Therefore, we amended our study population to exclude HPFS data but obtained the HPFS PI's agreement to be able to access to their data for cross validation. Based on the available samples, we utilized matched case control design to select the blood samples to be measured. The selected blood samples are currently being analyzed in the lab and we expect to receive the lab data by early 2015. While we are waiting for the lab results, we have been working on analyzing the related metabolic biomarkers and writing up manuscripts.					
<b>15. SUBJECT TERMS</b> Prostate cancer survivorship, metabolite profiling, metabolic biomarkers					
<b>16. SECURITY CLASSIFICATION OF:</b>			<b>17. LIMITATION OF ABSTRACT</b>	<b>18. NUMBER OF PAGES</b>	<b>19a. NAME OF RESPONSIBLE PERSON</b> USAMRMC
<b>a. REPORT</b>	<b>b. ABSTRACT</b>	<b>c. THIS PAGE</b>			<b>19b. TELEPHONE NUMBER</b> (include area code)
Unclassified	Unclassified	Unclassified	Unclassified		

Standard Form 298 (Rev. 8-98)  
Prescribed by ANSI Std. Z39.18

## Table of Contents

	<u>Page</u>
1. Introduction.....	4
2. Keywords .....	4
3. Accomplishments .....	4
4. Impact .....	10
5. Changes/Problems .....	11
6. Products .....	11
7. Participants & Other Collaborating Organizations .....	12
8. Special Reporting Requirements .....	13
9. Appendices .....	13

## 1. Introduction

A major challenge in prostate cancer (PCa) research is to distinguish aggressive from indolent disease. Although the D'Amico risk stratification is helpful and widely used to guide PCa treatment, it relies on a few standard clinical parameters (prostate specific antigen (PSA), stage, and grade) and cannot always reliably distinguish patients who will die from PCa from those who do not, leading to over-treatment and unnecessary side effects in many men with “low-risk” disease, preventing PCa-specific mortality only in a small minority. On the other hand, some cancers may be destined to recur despite aggressive multi-modality therapy. There is an urgent need for additional biologically relevant markers to improve prognostication beyond algorithms based solely on PSA, stage, and grade. Ideally, such biomarkers could also provide clinical guidance for alternative or novel treatments. Our prospective studies of adiposity, physical activity, and several individual biomarkers in two large Harvard cohorts demonstrate that markers of energy metabolism such as insulin, adipokines, and *de novel* fatty acid synthesis may play important roles in risk of lethal PCa. Recent development of a metabolite profiling platform by Dr. Clish's laboratory at the Broad Institute of MIT/Harvard further showed promising potential along this line of research. This technology has identified *in vitro* an aberrant activation of the PI3K downstream target as a common molecular event in cancer pathology and obesity; revealed significant associations of several amino acids and lipid metabolites in human plasma as signatures of insulin resistance and diabetes risk; and identified signatures of exercise performance and cardiovascular disease susceptibility, proving its validity of metabolic profiling. In addition, the methods have also passed our own rigorous reproducibility assessments. All these provide important ground work for the current proposal. The current study is using a targeted, LC-MS-based metabolite profiling platform to measure and compare metabolic profiles of prediagnostic blood samples collected from men subsequently diagnosed with PC and sample of men who remained cancer-free in Physicians' Health Study (PHS). And test whether these relationships are independent of the known metabolic risk factors (overweight/obese, insulin marker C-peptide, insulin-like growth factor I (IGF-I), IGF binding protein 3, (IGFBP-3), and adiponectin) as well as the clinical characteristics defined as the D'Amico risk.

## 2. Keywords

Prostate cancer survivorship, metabolomic profiling, metabolic biomarkers

## 3. Accomplishments

### What were the major goals of the project?

The three original aims were:

Aim 1: Explore and validate the metabolomic footprints for normal controls (n=50 x 2 cohorts) vs. three groups of cases (metastatic PC at diagnosis, initially localized PC and long-time survivors, and initially localized PC but died of PC; n=50 in each of the three groups, total n=150 cases x 2 cohorts);

Aim 2: Among men with initial localized PCa, explore and validate the metabolomic footprints for long-term survivors vs. men who subsequently died of PCa;

Aim 3: Test and validate whether these associations are independent of the known metabolic risk factors (overweight/obese, insulin marker C-peptide, insulin-like growth factor I (IGF-I), IGF

binding protein 3 (IGFBP-3), and adiponectin), as well as the clinical characteristics defined as the D'Amico risk.

In the original protocol, we plan to measure samples of PCa cases from both HPFS and PHS. However, the HPFS team has received separate grant for metabolomics measurement. Therefore, we amended our study population to exclude HPFS data, instead we use the fund to increase the sample size from 50 to 100 for each of the proposed aims so that the total sample size remain unchanged.

#### Modified aims:

Aim 1: 1a. Compare to 100 healthy men without cancer (at least at the time when the cases were diagnosed), metabolomic profiling for 100 men with "high risk" (T1-3 and Gleason 8+) or metastasis at diagnosis; 1b. Among men with "high risk" (T1-3 and Gleason 8+) or metastasis at diagnosis, metabolomic profiling between men who died of the cancer vs. those who were still alive by 2012;

Aim 2: Compare the metabolomic profiles between 100 men with "low-intermediate risk" (T1-3 and Gleason 2-7) PCa and died of the cancer with those (n=100) who survived at least 10 years after diagnosis.

Aim 3: Test whether these associations are independent of the known metabolic risk factors (overweight/obese, insulin marker C-peptide, insulin-like growth factor I (IGF-I), IGF binding protein 3 (IGFBP-3), and adiponectin), as well as the clinical characteristics defined as the D'Amico risk.

#### **What was accomplished under these goals?**

Based on the available samples, we utilized matched case control design to select the blood samples to be measured. The selected blood samples are currently being analyzed in the lab and we expect to receive the lab data by early 2015. While we are waiting for the lab results, we have been working on analyzing the related metabolic biomarkers and writing up manuscripts.

#### I. Sample selection

Based on the available samples, we utilized matched case control design to select the blood samples to be measured. The following part showed the matching method for each specific aim.

Aim 1: Study Population:

- 1) All population in the cohort;
- 2) Blood volume >100 ml;

Cases:

- 1) Incidence PCa cases
- 2) Localized and high grade cases or mets at diagnosis;
- 3) status in mortality file: died of PCa or alive;

Control Matching criteria:

- 1) Same age group at baseline: 40-<50 50-<60 6-<70 70+years
- 2) Same fasting  $\geq 8$ , 0<8 hrs
- 3) Controls have no cancer, or have cancer, but diagnosed after the last PCa diagnosis date in the age and fasting group.( we only have cancer information, no other disease info)

4) Frequency matching: get same percentile at each age group and total controls are 100.

Program: SAS proc surveysselect

Aim2: Study Population:

- 1) Incidence PCa cases with blood collected in 1982;
- 2) Localized and low grade cases;
- 3) Blood volume >100 ml;

Cases:

- 1) Status in mortality file: 1)PCa death; 2) Survived less than 10years

Control Matching criteria:

- 1) Same age at diagnosis
- 2) Same fasting  $\geq 8$ ,  $0 < 8$  hrs
- 3) Same Gleason category: 2-6,7
- 4) Controls are alive now, and alive more than 10 years.
- 5) 1:1 match", program: proc sql and hash table

In summary, for aim 1, we selected 100 cases who have localized, high grade PCa (clinical stage T1-T3, Gleason grade 8-10) or who have metastatic PCa (clinical stage T4N1M1, Gleason grade 2-10). We matched 100 controls from participants in PHS who were cancer free or have cancer, but diagnosed after the same group of cases. For aim 2, 48 eligible cases were identified from PHS, those cases have localized low grade (clinical stage T1-T3, Gleason grade 2-7) PCa who died of PCa within 10 years after diagnosis. Among them, 43 got matched with controls, who have localized low grade (clinical stage T1-T3, Gleason grade 2-7) PCa, alive at the end of this study or who have been alive more for than 10 years (Table 1).

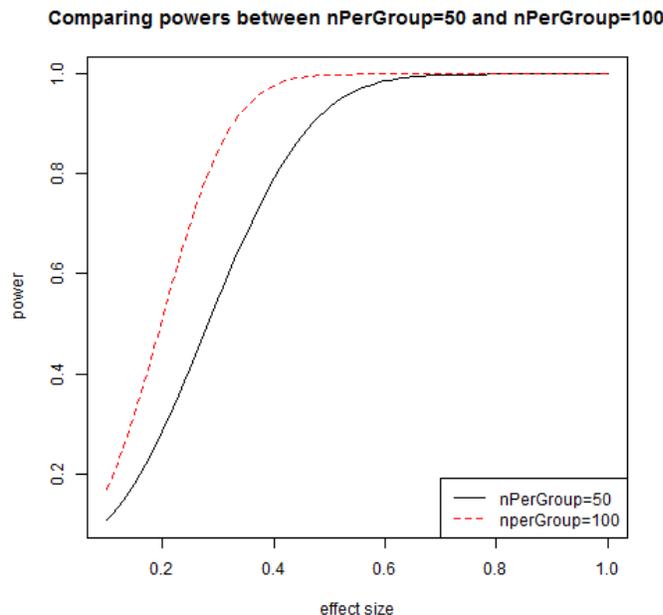
**Table 1: Final Result of sample selection**

Classification (N)		Sample size	Matched controls
<b>Aim 1 (All Population)</b>			
Localized T1-T3 & Gleason 8-10	Died of PCa	43	
	Alive	57	
Metastatic PCa (T4N1M1)	Died of PCa	51	
	Alive	30	
<b>Total</b>		<b>181</b>	
Healthy controls			
<b>Aim 2 (Incident PCa with 82 blood)</b>			
Localized T1-T3 & Gleason 2-6,7	died of PCa	<b>48</b>	
	Long-term survivor 10 yr+ controls	487	

Following this matching procedure, we selected a total of 372 (Aim1: 181 cases and 100 matched controls; Aim 2: 48 cases and 43 matched controls) blood samples for analyses. 28 QC samples were also included. After checking with the blood lab staff, we have only successfully identified 329 out of 372 eligible samples, and 40 quality control samples from lab. We then decided to add 31 additional samples from PCa patients who have been died from other cancers. We sent a final number of 400 samples to Dr. Clish's lab at November, 2014.

### Power consideration:

The data generated in Aims 1a, 1b, and 2 are all from matched/paired design. Hence, we use the same power calculation formula (Cohen, 1988) to estimate the powers given the sample size and type I error rate. We set the type I error rate as 0.05. The differences among the 3 aims are the effect size, which is unknown until we obtain the data. Hence, we tried a sequence value of effect size. In our original plan, we have  $n=50$  subjects for each of the 2 groups. In the current plan, we have  $n=100$  subjects per group. The power curves shown in the following Figure shows that the power is significantly improved when the sample size is increased from  $n=50$  per group to  $n=100$  subjects per group. For  $n=50$  per group, as long as the effect size  $> 0.41$ , the power would be  $> 0.8$ . For  $n=100$  per group, as long as the effect size  $> 0.29$ , the power would be  $> 0.8$ .



### Some concerns regarding the PI's access to the HPFS data for validation:

Because the HPFS have separate funding to measure metabolomics for their PCa cases and controls, we decided to focus this grant solely on the PHS to avoid redundant work but we will validate the model with the HPFS data even though the DOD is not paying for the analysis. To reassure the DOD that the PI could do so, we have a letter of agreement from Dr. Mucci, the co-leader of the HPFS SPORE project, to confirm that the PI Dr. Ma will be granted access to the HPFS data so that we can complete the validation.

## II. Completed/ongoing studies & results

### a. Insulin-like growth factor (IGF) pathway genetic polymorphisms, circulating IGF1 and IGFBP3 levels and prostate cancer survival

We conducted kernel machine pathway analysis to evaluate whether 530 tagging single-nucleotide polymorphism (SNP) in 26 IGF pathway-related genes were collectively associated with prostate cancer mortality among 5,887 prostate cancer patients (704 prostate cancer deaths) from 7 cohorts in the NCI Breast and Prostate Cancer Cohort Consortium (BPCa3).

IGF signaling pathway was associated with prostate cancer mortality ( $P=0.03$ ), and SNP sets of *IGF2-AS* and *SSTR2* were the main contributors (both  $P=0.04$ ) (Table 5). In SNP-specific analysis, 36 SNPs were associated with prostate cancer mortality with  $P_{\text{trend}} < 0.05$  but only 3 SNPs in the *IGF2-AS* remained significant after gene-based corrections. Two of the three SNPs

were in perfect linkage disequilibrium ( $r^2=1$  for rs1004446 and rs3741211) whereas the third rs4366464 was independent ( $r^2=0.03$ ). The hazard ratios (HRs) per each additional risk allele were 1.19 (95% CI 1.06-1.34;  $P_{\text{trend}}=0.003$ ) for rs3741211 and 1.44 (1.20-1.73;  $P_{\text{trend}}=0.0001$ ) for rs4366464. Rs4366464 remained significant after correction for all the SNPs tested ( $P_{\text{trend,corr}}=0.04$ ,  $M_{\text{eff}}=424$ ). Pre-diagnostic circulating levels of IGF1 (HR<sub>highest vs lowest quartile</sub> 0.71; 95%CI 0.48-1.04) and IGFBP3 (HR 0.93; 95%CI 0.65-1.34) were not associated with prostate cancer mortality.

The manuscript has been published by JNCI (June, 2014).

### **b.Characterization of energy-related biomarkers measured before and after PCa diagnosis in predicting all-cause and PCa-specific mortality.**

In the PHS, we defined “high energetic risk” as BMI>25 kg/m<sup>2</sup> and elevated C-peptide levels (in the highest quartile). We found that this “energetic risk” significantly predicted PCa mortality among men with localized disease at diagnosis independent of clinical characteristics. We replicated this association in an independent cohort, the Health Professionals Follow-up Study (HPFS).

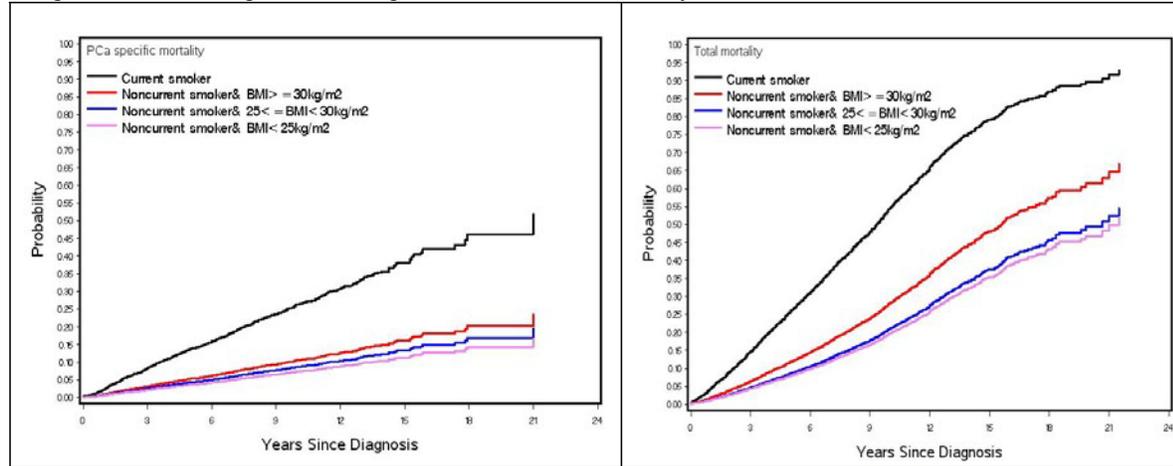
In both cohorts, we found that incorporating this “energetic risk” to the D’Amico risk score (defined by three clinical perimeters: PSA, clinical stage, and Gleason score) significantly improved the predictability of PCa-specific mortality and all-cause mortality in men with initial diagnosis of localized cancer; the C-statistic for PCa-specific mortality was improved from 0.72 to 0.78 ( $P<0.001$ ). Moreover, “energetic risk” identified ~20% of patients who are at high risk of disease specific mortality but are classified as low risk according to clinical characteristics. The resulting paper is undergoing peer-review. One major concern raised was the potential confounding by comorbidity and treatments. We therefore carefully evaluated the impact of these two factors from both cohorts and found little changes of the overall results.

### **c.Pre-diagnostic Obesity, Smoking and PCa survival**

Although obesity and smoking has not been strongly associated with prostate cancer (PCa) incidence, merging evidence linked them to increased PCa-specific mortality.we investigated the

associations of pre-diagnostic BMI and smoking status with risk of progression from time of PCa diagnosis to fatal outcome among 10,106 PCa patients from the NCI Breast and Prostate Cancer Cohort Consortium (BPC3).

Figure 1. Age-adjusted cumulative incidences of a) PCa-specific mortality; and b) Total mortality stratified by BMI categories and smoking status among PCa men from BPC3 study



Cumulative incidence curves show the probability of prostate cancer-specific mortality or total mortality after diagnosis according to baseline smoking and BMI categories, controlling for age at diagnosis.

The cumulative PCa-specific and overall mortality was much higher in current smokers as compared with never or former smokers. In contrast, the difference according to BMI categories among non-current smokers is much smaller, but still apparent for total mortality. This study provides further evidence that overweight/obesity and smoking history prior to diagnosis are related to poor survival among patients with PCa. The manuscript has been developed and is now circulating among coauthors.

#### d. Type 2 Diabetes before and after PCa diagnosis with PCa-specific and all-cause mortality.

Utilizing the same cohorts data from BPC3, we also observed that New T2D cases after PCa diagnosis was linked to improved survival among PCa cases (Table 2). We plan to look into this together with T2D related SNPs and C-peptide information.

Table 2: Diabetes status and prostate cancer/other mortality in the BPCa3 cohort

Variable	Prostate cancer specific mortality			Other mortality		
	HR	95%CI	p value	HR	95%CI	p value
Never	Ref			Ref		Ref
Diabetes status						
Before PCa diagnosis	1.02	0.95 1.09	0.583	1.25	1.08 1.46	0.004
After PCa diagnosis	0.72	0.61 0.84	<.001	0.62	0.41 0.93	0.021

Multivariate model adjusted for age at diagnosis (continuous), smoking status (never, former, current), BMI (18-22.9, 23-24.9, 25-27.9, 28-29.9, 30-34.9 kg/m<sup>2</sup>), drinking status (never, <15g/day, ≥ 15 & <30g/day, ≥30g/day), diabetes status (never, baseline, new), cohort (ATBC, CPS2, EPIC, HPFS, MCCS, MEC, PHS, PLCO), duration between baseline and PCa diagnosis (continuous);

#### e. GWAS-identified type 2 diabetes SNPs and risk of progression to fatal prostate cancer

This study will be based on our recently published study using the BPCa3 genome-wide association study of 2,782 advanced PCa cases and 4,458 controls to evaluate whether 36 T2D susceptibility loci and PCa incident risk (*M Machiela et al. Am J Epidemiol 2012*). Ten T2D markers near 9 loci (NOTCH2, ADCY5, JAZF1, CDKN2A/B, TCF7L2, KCNQ1, MTNR1B, FTO, and HNF1B) were nominally associated with PCa risk ( $P < 0.05$ ); the association for rs757210 at the HNF1B locus was significant when multiple comparisons were accounted for (adjusted  $P = 0.001$ ). Genetic risk scores weighted by the T2D log odds ratio and multilocus kernel tests also indicated a significant relation between T2D variants and PCa risk. These T2D risk variants have not been fully investigated for PCa progression to fatal outcome. Also, few studies have T2D phenotypes or sufficient power to assess whether T2D status mediates the relationship between T2D risk variants and PCa risk. We will fully evaluate these genes and mediation through or interaction with T2D for fatal PCa. We will also explore the association between T2D risk variants and risk of specific type of PCa cases (advanced PCa and died, advanced PCa and long term survivors 10+, localized PCa cases) For our analysis, all 36 T2D SNPs have been pulled out for 7,240 participants including both cases and controls from the imputed data files for BPCa3 Adv Prostate Cancer GWAS. The imputation was done using HapMap 2 Rel 22 CEU phased data as the reference panel.

**f. Elevation of circulating branched-chain amino acids is an early event in human pancreatic adenocarcinoma development.**

The pancreatic cancer group from Dana Farber Cancer Institute has also been working on the metabolomics of pancreatic cancer development. We are also working closely with this group on metabolomics data analysis.

This study utilized profiled metabolites in prediagnostic plasma from individuals with pancreatic cancer (cases) and matched controls from four prospective cohort studies. And find that elevated plasma levels of branched-chain amino acids (BCAAs) are associated with a greater than twofold increased risk of future pancreatic cancer diagnosis. This elevated risk was independent of known predisposing factors, with the strongest association observed among subjects with samples collected 2 to 5 years before diagnosis, when occult disease is probably present. We show that plasma BCAAs are also elevated in mice with early-stage pancreatic cancers driven by mutant *Kras* expression but not in mice with *Kras*-driven tumors in other tissues, and that breakdown of tissue protein accounts for the increase in plasma BCAAs that accompanies early-stage disease. Together, these findings suggest that increased whole-body protein breakdown is an early event in development of pancreatic ductal adenocarcinoma (PDAC).

This manuscript has been published in Nature Medicine (Sep, 2014).

**What opportunities for training and professional development has the project provided?**

This provided has provided funding and research opportunities for several doctoral and post-doctoral students from Harvard T.H. Chan School of Public Health.

Yin Cao, graduated from the doctoral of science program from Epidemiology department, and a current post-doc student at Nutrition Department. One of her thesis paper was based on and supported by the current project.

Changzheng Yuan, doctoral candidate at Nutrition and Epidemiology Department. She is now working on three research topics related to this project, mainly focusing on obesity, T2DM and genetic variants related to prostate cancer development.

Meng Yang, postdoc fellow at Nutrition Department. She currently working on studying the BMI trajectory, dietary factors, metabolic biomarkers and prostate cancer survivorship.

CY and MY also work closely with the project leader and statisticians to discuss the study design and sample selections.

**How were the results disseminated to communities of interest?**

Nothing yet to report.

**What do you plan to do during the next reporting period to accomplish the goals?**

We plan to conduct the analyses based on the proposed aims of metabolomic analysis after receiving the lab analyses results.

#### **4. Impact**

**What was the impact on the development of the principal discipline(s) of the project?**

The most notable strength of our proposal is the use of unbiased metabolomic profiling to distinguish lethal from indolent disease, a major challenge in prostate cancer research. Prostate cancer accounts for 25% of all newly diagnosed cancers and 9% of all cancer deaths in men, making it the most commonly diagnosed and second most lethal cancer for men in the United States. Wide spread use of PSA screening has changed the stage and grade distribution of disease at diagnosis but appears to have only modest effects on prostate cancer mortality. In the United States, 80 to 90% of prostate cancer cases are confined to the prostate and two-thirds of the cases are localized or regional disease and low- to moderate- grade at diagnosis. Current use of clinical features cannot always reliably distinguish patients who will die from prostate cancer from those who do not. Thus, it is important to identify novel markers specifically associated with lethal prostate cancer and lifestyle factors that influence disease progression.

The short-term (1-3 years) impact of this prospective study will be to provide a deeper understanding of the mechanisms of lethal PC phenotype so that relevant biological pathways can be revealed and new biomarkers can be developed. Findings from this study will help better understand the mechanisms of action of energy balance in tumor growth and metastasis and reveal novel biomarkers and pathway effects. This line of research is especially important in the context of highly prevalent obesity and hyperinsulinemia among nondiabetic U.S. adults in recent decades. Findings of this study could then provide biological rationale that risk of lethal PC could be reduced before or at early stage of the disease by modifying metabolic risk through physical activity, healthy diets, and other innovative approaches.

The long-term (3-8 years) goal of our biomarker study is to provide targeted patient identification and stratification to link patients with common biomarkers to the appropriate therapy. Our research could be extended to larger validation studies using novel biomarkers for better stratification methods than the current clinical parameters (e.g., D'Amico risk) to link

patients with common biomarkers to the appropriate personalized prevention and therapeutic strategies. The new risk stratifications could then help to identify candidates for randomized trials of novel agents targeting metabolic dysregulation and pathways. Ultimately, these novel biomarkers could also be candidates for response to such interventions.

The overarching challenges and focus areas of this proposal are discovery and validation of biomarkers for the detection and prediction of lethal prostate cancer. This study addresses the overarching challenge and one of the PY11PCR focus areas: discovery and validation of biomarkers for the detection and prediction of lethal prostate cancer from indolent disease, so that men with indolent disease could be spared from over-treatment, whereas those with high risk potential for lethal phenotype could receive appropriate personalized interventions at an early stage.

**What was the impact on other disciplines?**

Nothing yet to report.

**What was the impact on technology transfer?**

Nothing yet to report.

**What was the impact on society beyond science and technology?**

Nothing yet to report.

## **5. Changes/Problems**

**Changes in approach and reasons for change**

As mentioned above, we now only focus on the PHS cohort.

**Actual or anticipated problems or delays and actions or plans to resolve them**

Nothing yet to report.

**Changes that had a significant impact on expenditures**

Nothing yet to report.

**Significant changes in use or care of human subjects, vertebrate animals, biohazards, and/or select agents**

Nothing yet to report.

## **6. Products**

**Journal Publications:**

- 1) Machiela MJ, Lindström S, Allen NE, et al. *Association of type 2 diabetes susceptibility variants with advanced prostate cancer risk in the Breast and Prostate Cancer Cohort Consortium*. Am J Epidemiol. 2012 Dec 15;176(12):1121-9. PMID: 23193118
- 2) Song Y, Chavarro JE, Cao Y, et al. *Whole Milk Intake Is Associated with Prostate Cancer-Specific Mortality among U.S. Male Physicians*. J Nutr. 2013 Feb;143(2): PMID:23256145

- 3) Cao Y, Lindström S, Schumacher F. Et , et al. *Insulin-like growth factor (IGF) pathway genetic polymorphisms, circulating IGF1 and IGFBP3 levels and prostate cancer survival*. JNCI. 2014 Jun; 106(6): PMID: 24824313
- 4) Mayers J, Wu C, Clish C, et al. Elevation of circulating branched-chain amino acids is an early event in human pancreatic adenocarcinoma development. *Nature medicine*. 2014 Sep. 20(10): PMID: 25261994

**Other publications, conference papers, and presentations:**

Yuan C, Cao Y, Chavarro J, Lindström S ..., Ma J. *Prediagnostic body-mass index, smoking and prostate cancer survival in a multi-cohort consortium study*. A poster was presented at the Frontier of Cancer Prevention Research 2013 Conference in Washington DC, Nov. 2013.

**7.Participants & Other Collaborating Organizations**

**What individuals have worked on the project?**

Name	Jing Ma
Project Role:	PI
Researcher Identifier	0000-0002-9132-0741
Nearest person month worked	2.52
Contribution to Project	As the project PI, Dr. Ma has leaded the weekly meetings for project team members. She direct and is responsible for the overall study design and performance, report and manuscript preparation
Funding Support	R01CA141298 (Stampfer) – 0.6 Calendar Month R01CA137178 (Chan) – 0.24 Calendar Month W81XWH-11-1-0529 (Chavarro)- 0.24 Calendar Month U45 CA10006 (Hu)- 7.08 Calendar Month U01CA155340 (Han-Sub)- 0.12 Calendar Month

Name	Jorge Chavarro
Project Role:	Other Significant Contributor
Researcher Identifier	N/A
Nearest person month worked	No Measurable Effort
Contribution to Project	Dr. Chavarro works closely with Drs. Ma, Clish, and Qiu on data analysis, interpretation, and manuscript preparation
Funding Support	HHSN275201000020C (Hu)- 2.40 Calendar Month CA-10-006 (Hu)- 2.28 Calendar Month W81XWH-11-1-0529 (Chavarro)- 4.20 Calendar Month

Name	HayHaiyan Zhang
Project Role:	Data Manager
Researcher Identifier	N/A
Nearest person month worked	3
Contribution to Project	Zhang is responsible for managing the database for Physicians' Health Study biobank, case-control selection and preparation for the biospecimen pulling list preparation.

**Has there been a change in the active other support of the PD/PI(s) or senior/key personnel since the last reporting period?**

Nothing to Report.

**What other organizations were involved as partners?**

Clary Clish, sub-contract PI, conducting metabolic analysis, the Metabolite Profiling Platform, Broad Institute of MIT/Harvard. Dr. Clish is an expert in metabolic profiling assay development and validations, and oversee the assay development, measurement, and data annotation at his laboratory.

## 8. Special Reporting Requirements

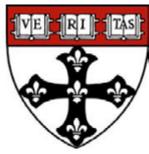
COLLABORATIVE AWARDS: For collaborative awards, independent reports are required from BOTH the Initiating PI and the Collaborating/Partnering PI. A duplicative report is acceptable; however, tasks shall be clearly marked with the responsible PI and research site. A report shall be submitted to <https://ers.amedd.army.mil> for each unique award.

QUAD CHARTS: If applicable, the Quad Chart (available on <https://www.usamraa.army.mil>) should be updated and submitted with attachments.

No.

## 9. Appendices

1. Machiela MJ, Lindström S, Allen NE, Haiman CA, Albanes D, Barricarte A, Berndt SI, Bueno-de-Mesquita HB, Chanock S, Gaziano JM, Gapstur SM, Giovannucci E, Henderson BE, Jacobs EJ, Kolonel LN, Krogh V, Ma J, Stampfer MJ, Stevens VL, Stram DO, Tjønneland A, Travis R, Willett WC, Hunter DJ, Le Marchand L, Kraft P. Association of type 2 diabetes susceptibility variants with advanced prostate cancer risk in the Breast and Prostate Cancer Cohort Consortium. *Am J Epidemiol.* 2012 Dec 15;176(12):1121-9. PMID: 23193118
2. Song Y, Chavarro JE, Cao Y, Qiu W, Mucci L, Sesso HD, Stampfer MJ, Giovannucci E, Pollak M, Liu S, Ma J. Whole Milk Intake Is Associated with Prostate Cancer-Specific Mortality among U.S. Male Physicians. *J Nutr.* 2013 Feb;143(2): PMID:23256145
3. Cao Y, Lindström S, Schumacher F, Stevens VL, Albanes D, Berndt SI, Boeing H, Bueno-de-Mesquita HB, Canzian F, Chamosa S, Chanock SJ, Diver WR, Gapstur SM, Gaziano JM, Giovannucci EL, Haiman CA, Henderson B, Johansson M, Le Marchand L, Palli D, Rosner B, Siddiq A, Stampfer MJ, Stram DO, Tamimi R, Travis RC, Trichopoulos D, Willett WC, Yeager M, Kraft P, Hsing AW, Pollak M, Lin X, Ma J. Insulin-like Growth Factor Pathway Genetic Polymorphisms, Circulating IGF1 and IGFBP3 Levels and Prostate Cancer Survival. *J Natl Cancer Inst.* 2014 May 13;106(6).
4. Mayers JR, Wu C, Clish CB, Kraft P, ....Ma J....Wolohin BM. Elevation of circulating branched-chain amino acids is an early event in human pancreatic adenocarcinoma development. *Nature Medicine* 2014 Oct. 20:1193-1198.
5. Yuan Y, Cao Y, Chavarro J, Lindström S, Qiu W, Drake B, Willett W, Hsing A, Kibel A, Rosner B, Stampfer M, Kraft P, Ma J. Prediagnostic body-mass index, smoking and prostate cancer survival in a multi-cohort consortium study. Manuscript circulated to co-authors.



# Prediagnostic Body-mass Index (BMI), Smoking and Prostate Cancer Survival



Changzheng Yuan<sup>1</sup>, Yin Cao<sup>1</sup>, Jorge Chavarro<sup>1,2</sup>, Sara Lindström<sup>2,3</sup>, Peter Kraft<sup>2,3</sup>, Jing Ma<sup>2,3</sup>  
(on behalf of the Breast and Prostate Cancer Cohort Consortium)

1. Department of Nutrition, Harvard School of Public Health, Boston, MA, 2. Channing Division of Network Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, 3. Department of Epidemiology, Harvard School of Public Health, Boston, MA

## BACKGROUND

- Meta-analysis linked elevated BMI with increased risk of PSA recurrence or prostate cancer (PC)-specific mortality. However, short follow-up and lack of control for smoking are major limitations in many of the clinical studies.
- Few prospective studies have sufficient power to investigate the relationship between obesity and lethal PC by time of BMI measurement before PC diagnosis and by smoking status.
- This study aimed to investigate the associations of pre-diagnostic BMI with risk of progression from time of PC diagnosis to fatal outcome, to study whether the relationship differs by time of BMI measurement before PC diagnosis and to assess effect modification by smoking status on BMI and PC survival.

## METHODS

- The study included 10,106 PC cases from the NCI Breast and Prostate Cancer Cohort Consortium (BPC3).
- BMI and smoking status were estimated at baseline before PC diagnosis. Deaths among PC patients were categorized into deaths from PC and other causes.
- We conducted the analysis in 3 parts according to exposures: BMI (18-22.9 kg/m<sup>2</sup>, 23-24.9 kg/m<sup>2</sup>, 25-27.9 kg/m<sup>2</sup>, 28-29.9 kg/m<sup>2</sup>, 30-34.9 kg/m<sup>2</sup>, 35 kg/m<sup>2</sup>+), smoking (Never, former, current) and their joint effect, respectively.
- Competing-risks regression model was used to take into account of other causes of death. Each analysis was performed on prostate cancer specific mortality and other mortality separately.

- Higher prediagnostic BMI was related to higher risk of dying from PC after diagnosis, the positive trend was mainly observed among men whose BMI measured >5 years before PC diagnosis.
- Smokers (both former and current) had significant higher risk of either PC specific mortality or other mortality, regardless of the time of measuring smoking status before PC diagnosis.
- The effect of BMI on PC survival is modified by smoking status. The positive trend of BMI with PC mortality was observed mainly among never and former smokers, but not among current smokers.

## RESULTS

Table 1. Characteristics of subjects according to death status

Variables	Total prostate cancer case (n=10,106)	Prostate cancer death (n=1,007)	Other death (n=1,886)	Survivors (n=7,213)
Age at diagnosis (mean±SD, yr)	68.7±6.7	69.5±7.2	71.7±6.3	67.8±6.5
BMI(mean±SD, kg/m <sup>2</sup> )	26.2±3.3	26.3±3.5	26.1±3.5	26.2±3.3
Duration between BMI and PCa diagnosis (mean±SD, yr)	7.3±4.7	6.9±4.3	7.4±4.6	7.3±4.7
Follow up time (mean±SD, yr)	8.2±3.9	5.1±3.6	6.5±3.9	9.1±3.6
Cohort (%)				
ATBC	10.2	28.6	24.2	4.0
CPS2	22.2	12.3	18.2	24.6
EPIC	15.3	20.8	9.3	16.1
HPFS	10.9	6.7	11.5	11.3
MCCS	9.5	5.7	7.0	10.6
MEC	6.5	2.8	6.6	7.0
PHS	13.5	17.2	14.7	12.6
PLCO	12.0	6.1	8.5	13.8
BMI category (%)				
18≤BMI<25	38.6	37.8	41.4	37.9
25≤BMI<30	49.0	49.3	45.8	49.8
30≤BMI	12.5	12.9	12.9	12.3
Smoking status (%)				
Never	36.3	23.3	24.7	41.2
Former	44.8	37.1	41.2	46.8
Current	18.9	39.5	34.2	12.1
Drinking status (%)				
Never	15.0	11.6	15.4	15.4
<15 g/day	52.2	54.0	51.6	52.0
≥15&<30 g/day	14.4	14.2	13.7	14.6
≥30 g/day	15.5	16.9	15.4	15.4
Diabetes status (%)				
Never	85.5	84.5	83.9	86.0
Baseline	4.1	4.7	6.5	3.3
New	7.2	3.1	5.8	8.1
Gleason score (%)				
G1(2-6)	40.4	11.7	33.7	46.2
G2(7)	26.4	19.2	20.8	28.9
G3(8-10)	11.4	22.5	12.5	9.6
Tumor stage (%)				
A/B (T1/T2)	81.5	61.7	79.1	84.9
C(T3)	10.3	13.5	8.6	10.3
D(T4)	3.8	18.4	3.0	2.0

Figure 1. Effect of BMI (smoking) on PC mortality and other mortality (overall and by measurement time before PC diagnosis)

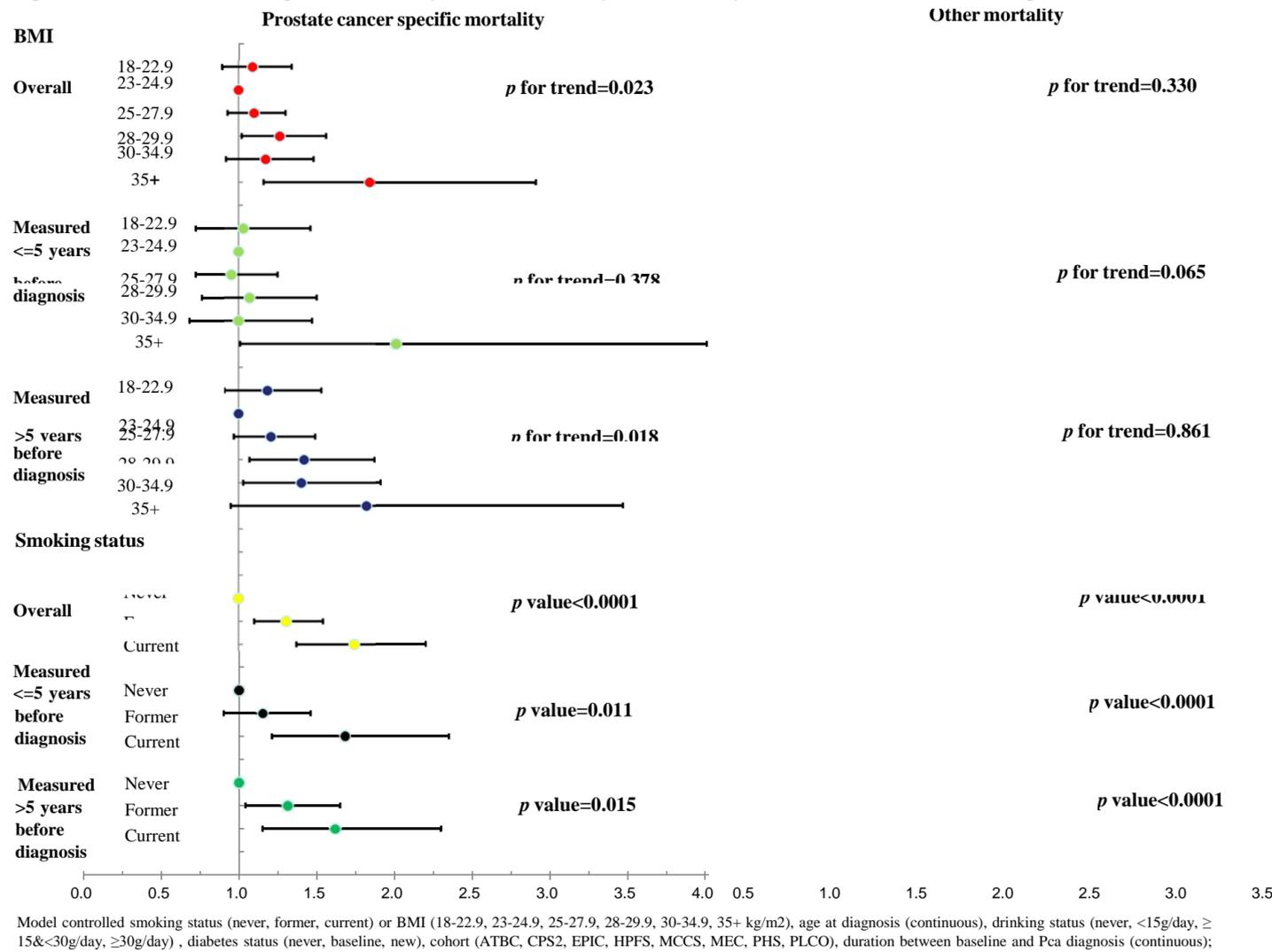
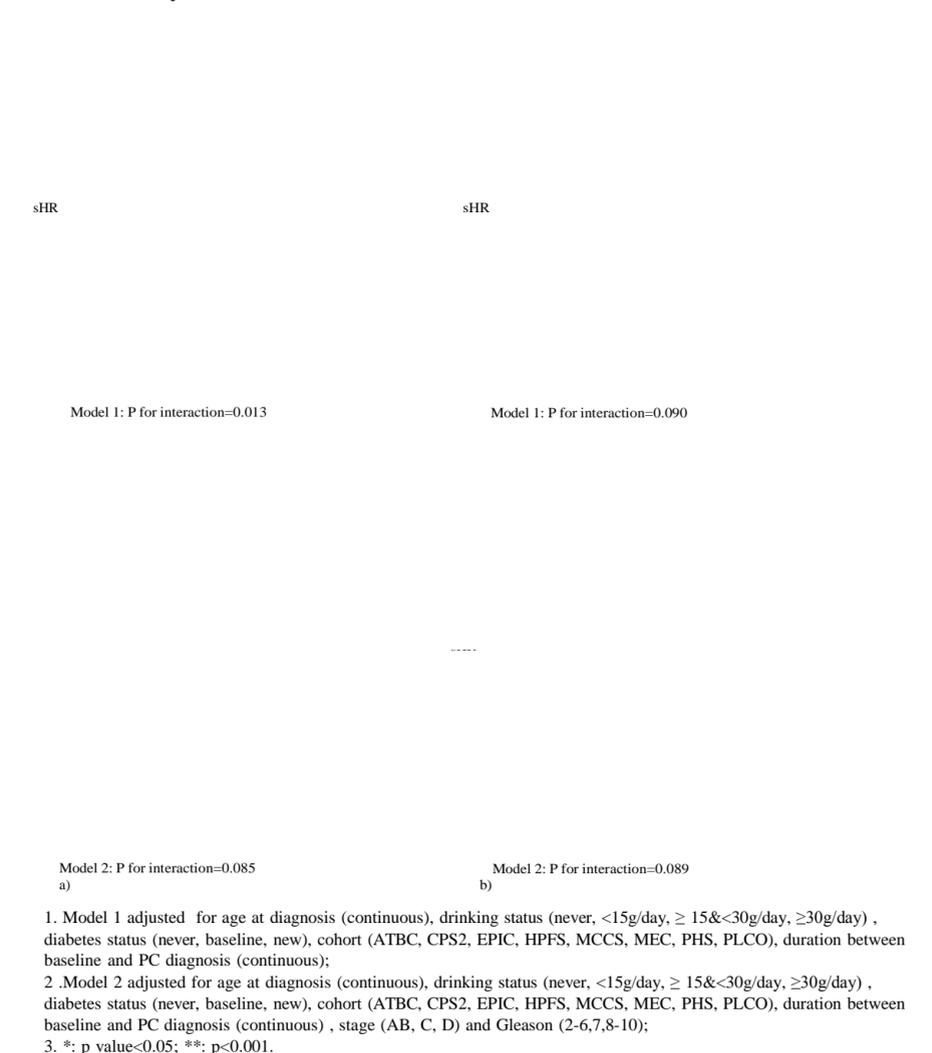


Figure 2. BMI and Smoking Joint Effects on a) Prostate Cancer Specific Mortality and b) Other mortality



# Elevation of circulating branched-chain amino acids is an early event in human pancreatic adenocarcinoma development

Jared R Mayers<sup>1,2,23</sup>, Chen Wu<sup>3–5,23</sup>, Clary B Clish<sup>6,23</sup>, Peter Kraft<sup>5,7</sup>, Margaret E Torrence<sup>1,2</sup>, Brian P Fiske<sup>1,2</sup>, Chen Yuan<sup>4</sup>, Ying Bao<sup>8</sup>, Mary K Townsend<sup>8</sup>, Shelley S Tworoger<sup>5,8</sup>, Shawn M Davidson<sup>1,2</sup>, Thales Papagiannakopoulos<sup>1,2</sup>, Annan Yang<sup>9</sup>, Talya L Dayton<sup>1,2</sup>, Shuji Ogino<sup>4,5,10</sup>, Meir J Stampfer<sup>5,8,11</sup>, Edward L Giovannucci<sup>5,8,11</sup>, Zhi Rong Qian<sup>4</sup>, Douglas A Rubinson<sup>4</sup>, Jing Ma<sup>5,8</sup>, Howard D Sesso<sup>5,12</sup>, John M Gaziano<sup>12,13</sup>, Barbara B Cochrane<sup>14</sup>, Simin Liu<sup>15,16</sup>, Jean Wactawski-Wende<sup>17</sup>, JoAnn E Manson<sup>5,8,12</sup>, Michael N Pollak<sup>18,19</sup>, Alec C Kimmelman<sup>9</sup>, Amanda Souza<sup>6</sup>, Kerry Pierce<sup>6</sup>, Thomas J Wang<sup>20</sup>, Robert E Gerszten<sup>6,21</sup>, Charles S Fuchs<sup>4,8</sup>, Matthew G Vander Heiden<sup>1,2,4,6</sup> & Brian M Wolpin<sup>4,22</sup>

Most patients with pancreatic ductal adenocarcinoma (PDAC) are diagnosed with advanced disease and survive less than 12 months<sup>1</sup>. PDAC has been linked with obesity and glucose intolerance<sup>2–4</sup>, but whether changes in circulating metabolites are associated with early cancer progression is unknown. To better understand metabolic derangements associated with early disease, we profiled metabolites in prediagnostic plasma from individuals with pancreatic cancer (cases) and matched controls from four prospective cohort studies. We find that elevated plasma levels of branched-chain amino acids (BCAAs) are associated with a greater than twofold increased risk of future pancreatic cancer diagnosis. This elevated risk was independent of known predisposing factors, with the strongest association observed among subjects with samples collected 2 to 5 years before diagnosis, when occult disease is probably present. We show that plasma BCAAs are also elevated in mice with early-stage pancreatic cancers driven by mutant *Kras* expression but not in mice with *Kras*-driven tumors in other tissues, and that breakdown of tissue protein accounts for the increase in plasma BCAAs that accompanies early-stage disease. Together, these findings suggest that increased whole-body protein breakdown is an early event in development of PDAC.

PDAC is a leading cause of cancer-related death worldwide, and most patients have incurable disease at diagnosis<sup>1</sup>. The best-characterized predisposing factors, current tobacco use and a first-degree relative with PDAC, both impart an approximate 1.8-fold increased risk for the disease<sup>5,6</sup>. These risk factors, however, have thus far provided limited insight into the biology of early disease progression of sporadic tumors. Development and progression of PDAC is also associated with altered systemic metabolism including obesity<sup>2</sup>, glucose intolerance<sup>3,4</sup> and cancer-induced cachexia<sup>7</sup>. Nevertheless, no systematic examination of circulating metabolites has been performed to determine whether altered metabolism may indicate subclinical pancreatic cancer or inform understanding of early disease progression when interventions might improve patient outcomes.

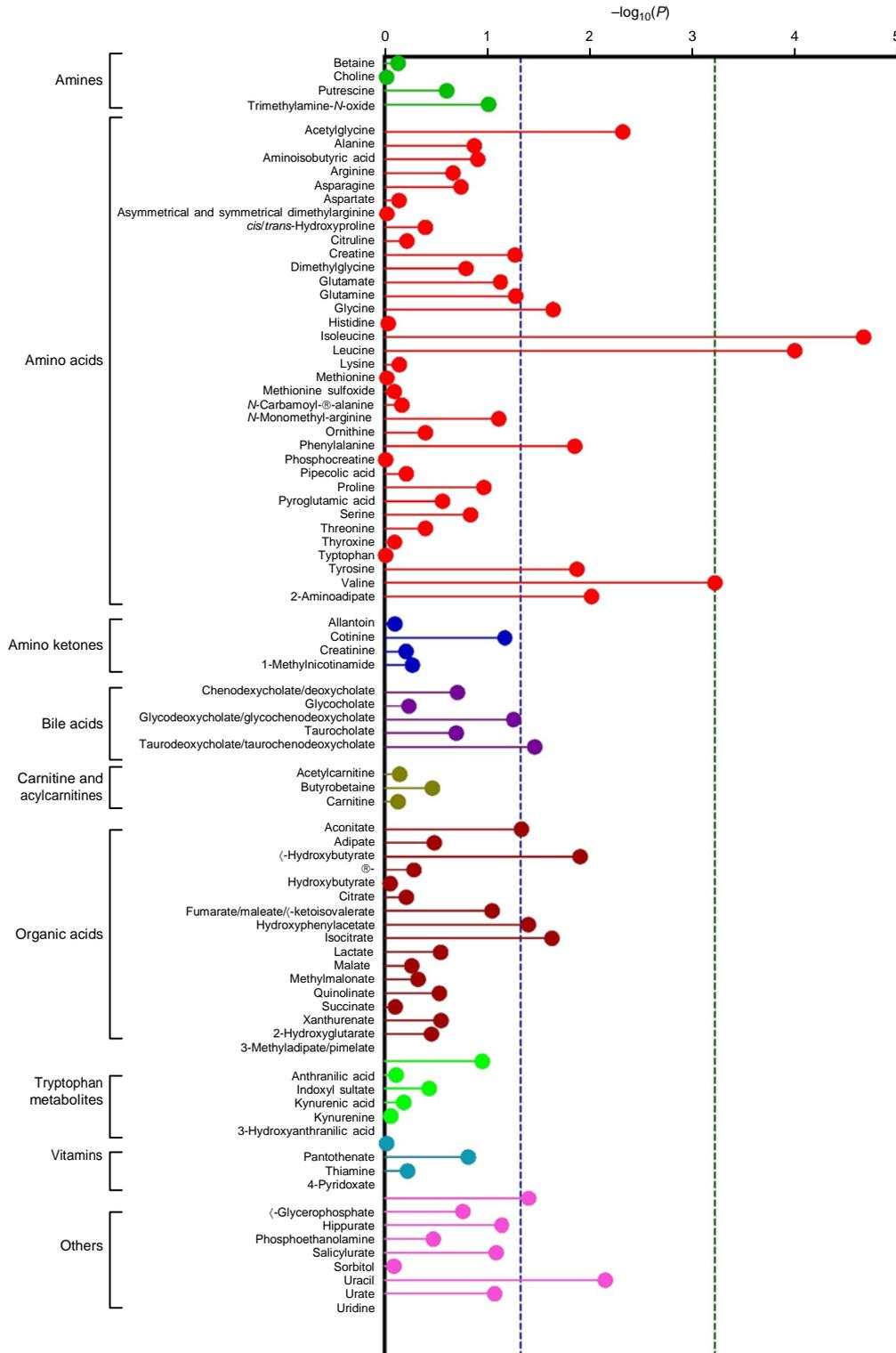
Prior efforts to identify changes in circulating metabolites related to cancer have employed a cross-sectional design, comparing cancer-free subjects to affected individuals with blood samples collected at diagnosis<sup>8–10</sup>. This approach is problematic for discovery of changes related to early cancer progression, as consequences of advanced disease are likely to have an impact on circulating metabolite profiles. This is particularly true for patients with pancreatic cancer, who commonly have significant anorexia, weight loss and pancreatic insufficiency at the time of diagnosis<sup>1</sup>. To investigate how altered metabolism might contribute to pancreatic malignancy, we profiled

<sup>1</sup>Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. <sup>2</sup>Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. <sup>3</sup>Department of Etiology and Carcinogenesis, Cancer Institute and Hospital, Chinese Academy of Medical Science and Peking Union Medical College, Beijing, China. <sup>4</sup>Department of Medical Oncology, Dana-Farber Cancer Institute and Harvard Medical School, Boston, Massachusetts, USA. <sup>5</sup>Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts, USA. <sup>6</sup>Broad Institute of MIT and Harvard University, Cambridge, Massachusetts, USA. <sup>7</sup>Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts, USA. <sup>8</sup>Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA. <sup>9</sup>Division of Genomic Stability and DNA Repair, Department of Radiation Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts, USA. <sup>10</sup>Department of Pathology, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA. <sup>11</sup>Department of Nutrition, Harvard School of Public Health, Boston, Massachusetts, USA. <sup>12</sup>Division of Preventive Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA. <sup>13</sup>Massachusetts Veterans Epidemiology Research and Information Center (MAVERIC), VA Boston Healthcare System, Boston, Massachusetts, USA. <sup>14</sup>University of Washington School of Nursing, Seattle, Washington, USA. <sup>15</sup>Department of Epidemiology, Brown University, Providence, Rhode Island, USA. <sup>16</sup>Department of Medicine, Brown University, Providence, Rhode Island, USA. <sup>17</sup>Department of Social and Preventive Medicine, University at Buffalo, State University of New York, Buffalo, New York, USA. <sup>18</sup>Department of Oncology, McGill University, Montreal, Quebec, Canada. <sup>19</sup>Department of Medicine, McGill University, Montreal, Quebec, Canada. <sup>20</sup>Division of Cardiovascular Medicine, Vanderbilt University, Nashville, Tennessee, USA. <sup>21</sup>Cardiology Division, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA. <sup>22</sup>Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA. <sup>23</sup>These authors contributed equally to this work. Correspondence should be addressed to M.G.V.H. (mvh@mit.edu) or B.M.W. (bwolpin@partners.org).

Received 26 June; accepted 5 August; published online 28 September 2014; doi:10.1038/nm.3686

plasma metabolites in cases with PDAC and matched controls drawn from four prospective cohort studies with blood collected at least 2 years before cancer diagnosis (**Supplementary Table 1**). The median time between blood collection and PDAC diagnosis was 8.7 years.

In conditional logistic regression models, levels of 15 metabolites were associated with future diagnosis of PDAC to  $P < 0.05$ ; three metabolites, the BCAAs isoleucine, leucine and valine were significant to  $P < 0.0006$ , the predefined significance threshold after correction



**Figure 1** Plasma metabolites and risk of future pancreatic cancer diagnosis.  $P$  values of the log-transformed, continuous metabolite levels from human plasma comparing pancreatic cancer cases and controls in conditional logistical regression models conditioned on matching factors and adjusted for age at blood draw (years, continuous), fasting time (<4 h, 4–8 h, 8–12 h,  $\geq 12$  h, missing) and race or ethnicity (white, black, other, missing). The dashed green line indicates the statistically significant  $P$  value threshold after Bonferroni correction for multiple-hypothesis testing,  $P < 0.0006$  (0.05/83). The dashed blue line indicates  $P$  of 0.05. The number of cases and controls analyzed for each metabolite is provided in **Supplementary Table 2**.

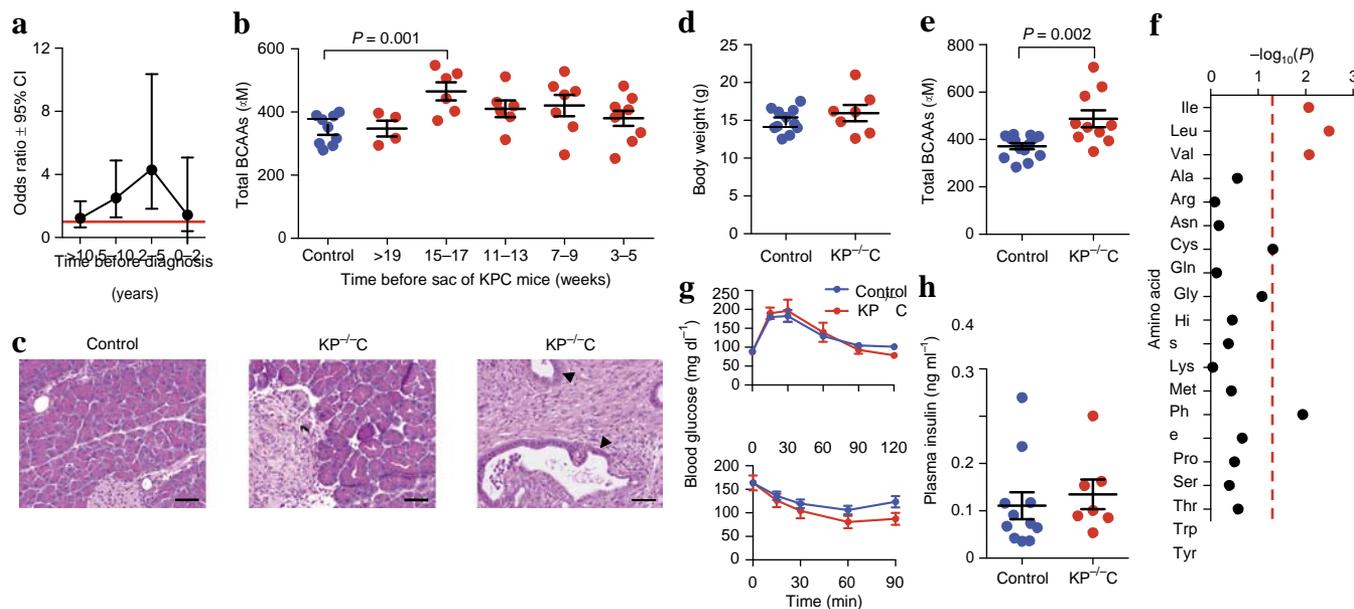
**Table 1 Odds ratios for pancreatic cancer by prediagnostic plasma levels of BCAAs**

Model <sup>a</sup>	Isoleucine		Leucine		Valine		Total BCAAs <sup>c</sup>	
	Extreme quintiles <sup>b</sup>	Per s.d.						
Base model	2.11 (1.40–3.18)	1.30 (1.15–1.48)	2.08 (1.38–3.13)	1.31 (1.14–1.50)	2.00 (1.37–2.92)	1.23 (1.09–1.39)	2.13 (1.43–3.15)	1.30 (1.14–1.48)
+ BMI and physical activity	2.05 (1.34–3.12)	1.29 (1.14–1.48)	2.01 (1.32–3.06)	1.29 (1.12–1.49)	1.94 (1.31–2.86)	1.21 (1.07–1.38)	2.06 (1.37–3.09)	1.28 (1.12–1.47)
+ BMI, physical activity and reported diabetes at blood collection	2.00 (1.31–3.05)	1.28 (1.13–1.46)	1.97 (1.29–2.99)	1.28 (1.11–1.48)	1.90 (1.28–2.81)	1.20 (1.06–1.37)	2.01 (1.34–3.03)	1.27 (1.11–1.46)
+ BMI, physical activity, reported diabetes, HbA1c, plasma insulin, proinsulin and C-peptide	1.86 (1.13–3.03)	1.24 (1.06–1.45)	1.81 (1.11–2.96)	1.24 (1.04–1.47)	1.67 (1.06–2.63)	1.14 (0.99–1.33)	1.89 (1.17–3.06)	1.22 (1.04–1.44)
Exclude subjects with reported diabetes or HbA1c $\geq 6.5\%$ at blood collection	2.12 (1.37–3.27)	1.33 (1.16–1.52)	2.16 (1.39–3.35)	1.32 (1.14–1.54)	1.91 (1.28–2.85)	1.23 (1.08–1.41)	2.19 (1.44–3.34)	1.31 (1.14–1.51)
Exclude subjects with reported diabetes or HbA1c $\geq 6.5\%$ at blood collection and those with reported diabetes after blood collection	2.18 (1.39–3.43)	1.32 (1.15–1.53)	2.21 (1.40–3.49)	1.33 (1.13–1.55)	1.94 (1.27–2.96)	1.22 (1.07–1.40)	2.25 (1.45–3.49)	1.31 (1.13–1.52)

<sup>a</sup>Odds ratio (95% CI) from conditional logistic regression models conditioned on matching factors and adjusted for age at blood draw (years, continuous), fasting time (<4 h, 4–8 h, 8–12 h,  $\geq 12$  h, missing) and race or ethnicity (white, black, other, missing). Subsequent models also adjusted for the indicated measure of energy balance, hyperglycemia or insulin resistance: body-mass index (kg/m<sup>2</sup>, continuous), physical activity (metabolic equivalent task-hours per week (MET-h/week), continuous), reported diabetes at blood collection (yes or no), hemoglobin A1c (HbA1c) (%), continuous, plasma insulin ( $\mu$ IU/ml, continuous), plasma proinsulin (pM, continuous) and plasma C-peptide (ng/ml, continuous). <sup>b</sup>Odds ratios (95% CI) for the comparison of the fifth quintile to the first quintile (referent) for the circulating BCAAs. <sup>c</sup>Sum of the concentrations of the three circulating BCAAs, isoleucine, leucine and valine.

for multiple-hypothesis testing (Fig. 1 and Supplementary Table 2). To evaluate the magnitude of risk for PDAC diagnosis, we divided participants into quintiles of increasing BCAA levels. Compared to

the bottom quintile, subjects in the top quintile had at least a two-fold increased risk of developing PDAC (Table 1, Supplementary Table 3 and Supplementary Fig. 1). As noted previously<sup>11</sup>, circulating levels



**Figure 2 Plasma BCAA levels are elevated during subclinical disease.** (a) Graph of odds ratio (error bars indicate 95% confidence interval (CI)) for future pancreatic cancer diagnosis among cohort cases and matched controls comparing highest versus lowest quintiles of circulating BCAA levels stratified by time from blood collection to the case's cancer diagnosis. Odds ratio was determined from conditional logistic regression models conditioned on matching factors and adjusted for age at blood draw (years, continuous), fasting time (<4 h, 4–8 h, 8–12 h,  $\geq 12$  h, missing) and race or ethnicity (white, black, other, missing). Red horizontal line marks an odds ratio of 1.0. The number of cases and controls in each time period and association results for the individual BCAAs are provided in Supplementary Table 6. (b) Graph of mean ( $\pm$ s.e.m.) total plasma BCAA concentration in *LSL-Kras<sup>G12D/+</sup>; LSL-Trp53<sup>R172H/+</sup>; Pdx1-cre* (KPC) mice over time and in littermate controls lacking either *LSL-Kras<sup>G12D</sup>* or *Pdx1-cre* or both. Each control data point is an average for one mouse over the course of the study (Supplementary Fig. 4b), and values for KPC mice living longer than 19 weeks are averaged for the >19-weeks time point. For weeks 15–17,  $n = 6$  KPC and  $n = 9$  control, Student's *t*-test,  $P = 0.001$ . For >19 weeks,  $n = 4$ ; 11–13 weeks,  $n = 6$ ; 7–9 weeks,  $n = 7$ ; 3–5 weeks,  $n = 9$ . (c) H&E staining of pancreatic tissue obtained from *KP<sup>-/-</sup>* mice and littermate controls at 3–4 weeks of age. Tissues are from a control mouse with histologically normal pancreas (left) a *KP<sup>-/-</sup>* mouse with areas of PDAC adjacent to areas of normal pancreas (middle) and a *KP<sup>-/-</sup>* mouse with areas of PDAC and pancreatic intraepithelial neoplasia (arrowheads; right). Scale bars, 50  $\mu$ m. (d) Mean ( $\pm$ s.e.m.) body weights at 3–4 weeks of age for *KP<sup>-/-</sup>* mice and littermate controls ( $n = 7$  *KP<sup>-/-</sup>*,  $n = 11$  control). (e) Mean ( $\pm$ s.e.m.) total plasma BCAA levels from *KP<sup>-/-</sup>* mice and littermate controls at 3–4 weeks of age ( $n = 10$  *KP<sup>-/-</sup>*,  $n = 14$  control, Student's *t*-test,  $P = 0.002$ ). (f)  $P$  values for comparison of circulating amino acid levels in *KP<sup>-/-</sup>* mice and littermate controls at 3–4 weeks of age ( $n = 10$  *KP<sup>-/-</sup>*,  $n = 14$  control). Red points indicate BCAAs. The dashed red line indicates  $P$  value of 0.05. (g) Top, glucose tolerance test in *KP<sup>-/-</sup>* mice and littermate controls at the time of weaning ( $n = 7$  *KP<sup>-/-</sup>*,  $n = 11$  control). Bottom, insulin tolerance test in *KP<sup>-/-</sup>* mice and littermate controls at 4 weeks of age ( $n = 7$  *KP<sup>-/-</sup>*,  $n = 15$  control). Error bars indicate s.e.m. (h) Mean ( $\pm$ s.e.m.) fasting plasma insulin levels from *KP<sup>-/-</sup>* mice and littermate controls at 4 weeks of age ( $n = 7$  *KP<sup>-/-</sup>*,  $n = 11$  control).

of the three BCAAs were highly correlated (Supplementary Table 4), reflecting their common pathways of metabolism<sup>12</sup> and leading to similar results for the sum total of BCAAs (Table 1 and Supplementary Table 3).

Circulating BCAAs are elevated in obese individuals and those with insulin resistance<sup>13</sup>. In study participants, plasma BCAA levels modestly correlated with markers of energy balance, obesity and glucose intolerance (Supplementary Table 4). To evaluate the independent effect of BCAAs on PDAC risk, we assessed models that included these markers and found that the odds ratios for PDAC remained largely unchanged (Table 1). Elevated circulating levels of BCAAs are also associated with future risk of diabetes<sup>11,14</sup>. As type 2 diabetes is a predisposing factor for PDAC<sup>15</sup>, we questioned whether the intermediate development of diabetes underlied the association of BCAAs with future PDAC diagnosis. Exclusion of subjects with diabetes at blood collection did not change our results (Table 1), indicating that we had not identified a signature of prevalent diabetes associated with later PDAC diagnosis. To determine whether increased circulating BCAAs identify a population at risk for diabetes, who are then at elevated risk of PDAC, we excluded subjects who developed diabetes between the time of blood collection and cancer diagnosis and found the results unchanged (Table 1). These data suggest that the association of circulating BCAAs with future PDAC diagnosis is not dependent on intermediate development of diabetes.

To examine the contribution of circulating BCAAs to risk stratification models for PDAC, we evaluated the area under the curve (AUC) of receiver-operating-characteristic (ROC) curves<sup>16</sup> and net reclassification improvement (NRI)<sup>17</sup> with low-risk and high-risk categories. Compared to the base model, including circulating BCAAs led to a significant increase in AUC (Supplementary Table 5a and Supplementary Fig. 2) and a net 8.2% of cases moving to the high-risk category with an NRI of 5% (Supplementary Table 5b). Thus, in our population, inclusion of circulating BCAAs in risk stratification models improved the ability to identify future PDAC cases.

In stratified analyses, we noted no significant differences in the association of BCAAs with PDAC by cohort, sex, smoking status,

body-mass index (BMI) or fasting status at blood collection (Supplementary Fig. 3, all interaction  $P \geq 0.14$ ). To examine when circulating BCAAs were most associated with PDAC, we stratified cases and matched controls by time interval between blood collection and PDAC diagnosis (<2 years, 2 to <5 years, 5 to <10 years and  $\geq 10$  years). These analyses demonstrated particularly strong associations between elevated BCAAs at 2–5 years before diagnosis and future PDAC diagnosis (Fig. 2a and Supplementary Table 6).

Experimental studies indicate years elapse between formation of the initial malignant clone and cancer diagnosis<sup>18</sup>, suggesting that occult PDAC might have been present at the time points showing the strongest associations with elevated BCAAs. We therefore hypothesized that elevated circulating BCAAs are a marker of early PDAC. To test this possibility, we conducted a prospective serial blood sampling study using lox-stop-lox (*LSL*)-*Kras*<sup>G12D/+</sup>; *LSL*-*Trp53*<sup>R172H/+</sup>; *Pdx1-cre* (KPC) mice, which develop PDAC with variable latency<sup>19</sup>. KPC mice progress through all histological stages of disease, from normal pancreata to invasive adenocarcinoma, with a median survival of approximately 21 weeks<sup>19</sup> (Supplementary Fig. 4a). KPC mice initially displayed similar BCAA levels to littermate controls, but they developed significant elevations from 15–17 weeks before death (Fig. 2b and Supplementary Fig. 4b). These data suggest circulating BCAA elevations accompany early PDAC.

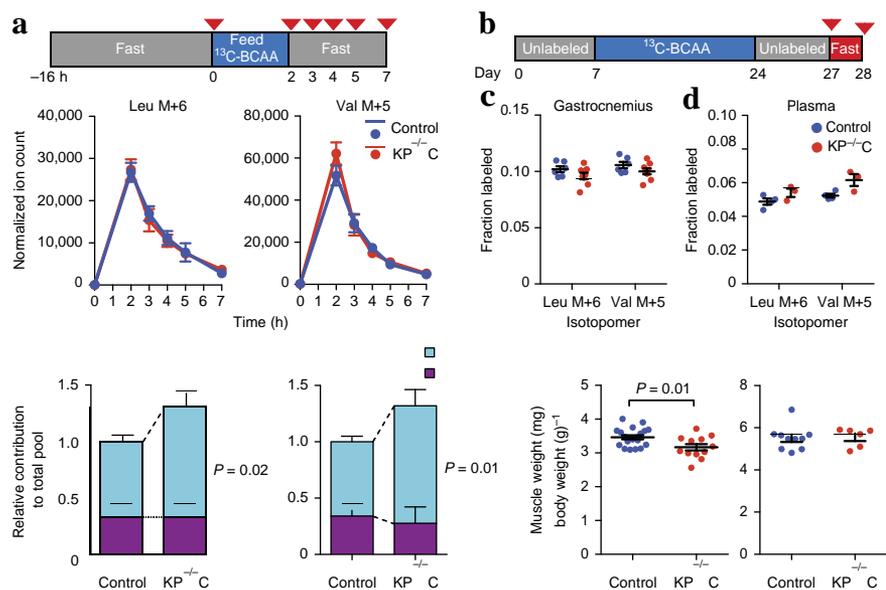
*LSL*-*Kras*<sup>G12D/+</sup>; *Trp53*<sup>flx/flx</sup>; *Pdx1-cre* (KP<sup>-/-</sup>C) mice develop PDAC with more consistent kinetics, displaying precursor lesions with limited invasive cancer by 3–4 weeks of age (Fig. 2c) and a median lifespan of 10–12 weeks<sup>20</sup>. In mice at 3–4 weeks of age, we observed no difference in body weight or food consumption between KP<sup>-/-</sup>C mice and littermate controls lacking either *Pdx1-Cre* or *LSL-Kras*<sup>G12D</sup> or both, suggesting animals with early PDAC had not yet developed overt constitutional symptoms (Fig. 2d and Supplementary Fig. 4c). Consistent with findings in patients and KPC mice, circulating BCAA levels were higher in KP<sup>-/-</sup>C animals with subclinical PDAC when compared with those in littermate control mice (Fig. 2e), a pattern not observed for most other amino acids (Fig. 2f and Supplementary Fig. 4d,e) We observed no significant differences in fasting blood

**Figure 3** BCAA elevations are derived from a long-term pool of amino acids. (a) Plasma levels (mean  $\pm$  s.e.m.) of <sup>13</sup>C-labeled leucine (M+6) and <sup>13</sup>C-labeled valine (M+5) normalized to food intake over time following a two-hour exposure to diets containing <sup>13</sup>C-labeled leucine and <sup>13</sup>C-labeled valine ( $n = 8$  KP<sup>-/-</sup>C,  $n = 6$  control). The time points correspond to the red arrowheads in the diagram. (b) Diagram of experiment using labeled diets to investigate contributions to plasma BCAA levels from long-term pools. Two cohorts of mice were used for these experiments, one killed in the fed state

time points indicated by the red arrowheads.

total amino acids in protein hydrolysate of gastrocnemius muscle from fasted KP<sup>-/-</sup>C mice and control littermates ( $n = 8$  KP<sup>-/-</sup>C,  $n = 6$  control). (d) Mean ( $\pm$ s.e.m.) fractional labeling of plasma amino acids in KP<sup>-/-</sup>C and control mice in the fed state ( $n = 3$  KP<sup>-/-</sup>C,  $n = 4$  control). (e) The calculated contribution ( $\pm$ s.e.m.) of the short- and long-term BCAA

pools to the BCAAs present in plasma ( $n = 22$  KP<sup>-/-</sup>C,  $n = 24$  control, two-sample *t*-test,  $P = 0.02$  for leucine,  $P = 0.01$  for valine). (f) Mean ( $\pm$ s.e.m.) gastrocnemius (a predominantly fast-twitch muscle) weight (left, Student's *t*-test,  $P = 0.01$ ), and heart weight (right) normalized to body weight of KP<sup>-/-</sup>C mice and littermate controls at 4 weeks of age ( $n = 6$  KP<sup>-/-</sup>C,  $n = 10$  control).



glucose, response to glucose load, response to insulin challenge or fasting insulin levels during intraperitoneal glucose and insulin tolerance tests in 4-week-old  $KP^{-/-}$ C and control mice (Fig. 2g,h and Supplementary 4f–i). These findings argue that BCAA elevations are not reflective of hyperglycemia or insulin resistance and instead represent an early consequence of subclinical PDAC.

We examined whether malignancies in other tissues induced by the same genetic lesions could cause elevated plasma BCAA levels. Cre recombinase introduction into lung or muscle of mice with the *LSL-Kras*<sup>G12D</sup> and *Trp53*<sup>fllox/fllox</sup> alleles from the  $KP^{-/-}$ C model leads to non-small-cell lung cancer and sarcoma, respectively<sup>21–23</sup>. Neither model displayed the BCAA alterations seen with early PDAC (Supplementary Fig. 5). Subcutaneous and orthotopic implantation of cancer cell lines derived from the  $KP^{-/-}$ C model into immunocompetent syngenic hosts both also failed to cause elevated BCAA levels (Supplementary Fig. 6). These data argue that elevations in BCAAs are associated with early-stage autochthonous tumors arising in the pancreas and are not a general feature of *Kras*-driven cancer. They also suggest that implantation of cells from end-stage disease does not model the early disease state that results in BCAA elevations.

Chronic pancreatitis is a risk factor for human PDAC<sup>24</sup>, and pancreatic inflammation can promote PDAC development and progression in mice<sup>25,26</sup>. Therefore, we examined whether BCAA elevations might be a cause or consequence of pancreatic inflammation in early disease in mice. Mild, chronic pancreatitis induced by caerulein in the absence of tumorigenesis failed to cause elevations in BCAAs (Supplementary Fig. 7a–h), and prolonged increases in plasma BCAA levels caused by dietary interventions did not cause pancreatic inflammation or pancreatitis (Supplementary Fig. 7i–o). Nevertheless, further studies are needed to understand the relationship between BCAAs and more severe pancreatitis.

Unlike levels of other amino acids, plasma BCAA levels are not regulated by the liver<sup>27,28</sup>; instead, levels are determined by dietary uptake, tissue utilization and breakdown of muscle and other body protein stores<sup>27,29</sup>. Therefore, plasma BCAAs may originate from short-term pools related to dietary uptake and disposal or long-term pools related to breakdown of tissue proteins. To determine the involvement of the short-term pool, we fed 4-week-old  $KP^{-/-}$ C mice and littermate control mice lacking either *Pdx1-Cre* or *LSL-Kras*<sup>G12D</sup> or both a defined amino acid diet, in which 20% of leucine and valine were <sup>13</sup>C labeled.  $KP^{-/-}$ C and control mice consumed similar amounts of food when exposed to labeled diet for 2 h (Supplementary Fig. 8a), and we observed no difference in appearance and disappearance of plasma <sup>13</sup>C label (Fig. 3a and Supplementary Fig. 8b), arguing that gut uptake and peripheral disposal of BCAAs are similar in mice with or without PDAC.

To determine the contribution of long-term BCAA pools to plasma levels, we exposed mice to labeled diet during a period of rapid growth in early life and then switched them to unlabeled diet for 3 d to chase label from the short-term pool (Fig. 3b). Despite similar peripheral tissue protein labeling (Fig. 3c and Supplementary Fig. 8c), the fraction of labeled BCAAs in plasma was elevated in 4-week-old  $KP^{-/-}$ C mice relative to that in littermate controls (Fig. 3d). Furthermore, by comparing the amount of label in plasma under fed conditions, encompassing both labeled long-term and unlabeled short-term pools, to that under fasted conditions, in which only labeled long-term pools contribute, we calculated that increased liberation of BCAAs from long-term body stores was solely responsible for the elevations in BCAAs in  $KP^{-/-}$ C mice (Fig. 3d,e and Supplementary Fig. 8d). These data suggest that an early consequence of PDAC is enhanced

breakdown of tissue proteins leading to elevated plasma BCAA levels. Consistent with this hypothesis,  $KP^{-/-}$ C mice with early PDAC had smaller fast-twitch muscles with no changes in slow-twitch and cardiac muscle weight (Fig. 3f and Supplementary Fig. 9). Notably, muscle atrophy associated with prolonged fasting and late-stage malignancy exhibits a similar pattern<sup>30–32</sup>.

Increased muscle catabolism represents one aspect of cancer-associated cachexia, a wasting syndrome frequently affecting patients with advanced PDAC and contributing to worse outcomes<sup>33–36</sup>. Our findings, however, suggest that protein breakdown begins much earlier than previously appreciated and predates onset of clinical cachexia. Inflammatory cytokines produced by immune and/or tumor cells have been implicated in cachexia<sup>31,37</sup>, and the low disease burden at the time of BCAA elevation suggests hormonal factors may be involved in early PDAC to cause these elevations as well. Liberation of tissue amino acids could support the elevated amino acid requirements of pancreatic cancer cells<sup>38,39</sup>, with BCAAs and/or other amino acids derived from tissue breakdown contributing to disease progression. Because hepatic metabolism maintains relatively constant plasma levels of all amino acids except BCAAs<sup>27,29,40</sup>, increased liberation of tissue amino acids would be expected raise BCAA concentrations in blood. The association between elevated BCAA levels and other metabolic disease states<sup>11,13,14,41</sup> suggests that high plasma BCAA concentrations could be a general marker of increased protein turnover, and elevated BCAA levels could contribute to the peridiagnostic hyperglycemia commonly found in patients with PDAC<sup>42</sup>.

In participants from four large prospective cohorts, circulating BCAAs were associated with future diagnosis of PDAC. We observed similar BCAA elevations in two mouse models of PDAC and demonstrated that these elevations result from breakdown of peripheral protein stores. These findings provide new insight into how early disease affects whole-body metabolism and suggest that muscle protein loss occurs much earlier in disease progression than previously appreciated.

## METHODS

Methods and any associated references are available in the [online version of the paper](#).

*Note: Any Supplementary Information and Source Data files are available in the online version of the paper.*

## ACKNOWLEDGMENTS

We would like to acknowledge A. Deik and K. Bullock of the Broad Institute for assistance with LC-MS sample analyses and the Tang Histology facility in the Koch Institute Swanson Biotechnology Center for assistance processing mouse tissues. Cambridge Isotope Laboratories supplied <sup>13</sup>C-BCAA diets for mouse labeling studies. We would like to thank C. Newgard and A. Goldberg for their thoughtful discussions regarding this manuscript. We would also like to thank the participants and staff of the Health Professionals Follow-Up Study (HPFS), Nurses' Health Study (NHS), Physicians' Health Study I (PHS) and Women's Health Initiative-Observational Study (WHI-OS) for their contributions as well as the cancer registries of the following states for their help: Alabama, Arizona, Arkansas, California, Colorado, Connecticut, Delaware, Florida, Georgia, Indiana, Illinois, Iowa, Kentucky, Louisiana, Maine, Maryland, Massachusetts, Michigan, Nebraska, New Hampshire, New Jersey, New York, North Carolina, North Dakota, Ohio, Oklahoma, Oregon, Pennsylvania, Rhode Island, South Carolina, Tennessee, Texas, Virginia, Washington and Wyoming. NHS and HPFS are supported by US National Institutes of Health (NIH) grants P01 CA87969, P01 CA55075, P50 CA127003, R01 CA124908, R01 CA49449 and 1UM1 CA167552. PHS is supported by NIH grants CA 97193, CA 34944, CA 40360, HL 26490 and HL 34595. The WHI program is funded by the NIH through contracts HHSN268201100046C, HHSN268201100001C, HHSN268201100002C, HHSN268201100003C, HHSN268201100004C and HHSN271201100004C. We acknowledge additional support from grant F30 CA183474 to J.R.M.; from a Nestle Research Center award to the Broad Institute; from R01 DK081572 grant to T.J.W. and R.E.G.; from the

Robert T. and Judith B. Hale Fund for Pancreatic Cancer, Perry S. Levy Fund for Gastrointestinal Cancer Research and Pappas Family Research Fund for Pancreatic Cancer to C.S.F.; from the Burroughs Wellcome Fund, Damon Runyon Cancer Research Foundation, the Smith Family and the Stern Family to M.G.V.H.; and from NIH/NCI grant K07 CA140790, the American Society of Clinical Oncology Conquer Cancer Foundation, the Howard Hughes Medical Institute and Promises for Purple to B.M.W. M.G.V.H. additionally acknowledges support from P30-CA14051 and P01-CA117969, and major support for this project was provided by the Howard Hughes Medical Institute to B.M.W. and the Lustgarten Foundation to C.S.F., M.G.V.H. and B.M.W.

## AUTHOR CONTRIBUTIONS

All authors participated in the analysis and interpretation of data. C.W., C.B.C., P.K., C.Y., Y.B., M.K.T. and B.M.W. performed the statistical analyses of the human data. C.B.C., M.K.T., S.S.T., T.J.W., R.E.G. and B.M.W. evaluated the platform for analysis of plasma metabolites in large cohort studies. C.B.C., J.R.M., A.S. and K.P. performed metabolite profiling experiments. P.K., Y.B., M.K.T., S.S.T., S.O., M.J.S., E.L.G., Z.R.Q., D.A.R., J.M., H.D.S., J.M.G., B.B.C., S.L., J.W.-W., J.E.M., M.N.P. and C.S.F. assisted in data acquisition, management and interpretation from the four cohort studies. J.R.M. and M.E.T. conducted all mouse experiments with assistance from B.P.F., S.M.D., T.P., A.Y., T.L.D. and A.C.K. C.W., C.B.C., J.R.M., C.S.F., M.G.V.H. and B.M.W. designed the study and drafted the manuscript with input from all authors.

## COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

- Vincent, A., Herman, J., Schulick, R., Hruban, R.H. & Goggins, M. Pancreatic cancer. *Lancet* **378**, 607–620 (2011).
- Michaud, D.S. *et al.* Physical activity, obesity, height, and the risk of pancreatic cancer. *J. Am. Med. Assoc.* **286**, 921–929 (2001).
- Stolzenberg-Solomon, R.Z. *et al.* Insulin, glucose, insulin resistance, and pancreatic cancer in male smokers. *J. Am. Med. Assoc.* **294**, 2872–2878 (2005).
- Wolpin, B.M. *et al.* Hyperglycemia, insulin resistance, impaired pancreatic beta-cell function, and risk of pancreatic cancer. *J. Natl. Cancer Inst.* **105**, 1027–1035 (2013).
- Lynch, S.M. *et al.* Cigarette smoking and pancreatic cancer: a pooled analysis from the pancreatic cancer cohort consortium. *Am. J. Epidemiol.* **170**, 403–413 (2009).
- Klein, A.P. Genetic susceptibility to pancreatic cancer. *Mol. Carcinog.* **51**, 14–24 (2012).
- Tan, B.H., Birdsell, L.A., Martin, L., Baracos, V.E. & Fearon, K.C. Sarcopenia in an overweight or obese patient is an adverse prognostic factor in pancreatic cancer. *Clin. Cancer Res.* **15**, 6973–6979 (2009).
- Claudino, W.M., Goncalves, P.H., di Leo, A., Philip, P.A. & Sarkar, F.H. Metabolomics in cancer: a bench-to-bedside intersection. *Crit. Rev. Oncol. Hematol.* **84**, 1–7 (2012).
- Sreekumar, A. *et al.* Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression. *Nature* **457**, 910–914 (2009).
- Kobayashi, T. *et al.* A novel serum metabolomics-based diagnostic approach to pancreatic cancer. *Cancer Epidemiol. Biomarkers Prev.* **22**, 571–579 (2013).
- Wang, T.J. *et al.* Metabolite profiles and the risk of developing diabetes. *Nat. Med.* **17**, 448–453 (2011).
- Brosnan, J.T. & Brosnan, M.E. Branched-chain amino acids: enzyme and substrate regulation. *J. Nutr.* **136**, 207S–211S (2006).
- Newgard, C.B. *et al.* A branched-chain amino acid-related metabolic signature that differentiates obese and lean humans and contributes to insulin resistance. *Cell Metab.* **9**, 311–326 (2009).
- Floegel, A. *et al.* Identification of serum metabolites associated with risk of type 2 diabetes using a targeted metabolomic approach. *Diabetes* **62**, 639–648 (2013).
- Huxley, R., Ansary-Moghaddam, A., Berrington de Gonzalez, A., Barzi, F. & Woodward, M. Type-II diabetes and pancreatic cancer: a meta-analysis of 36 studies. *Br. J. Cancer* **92**, 2076–2083 (2005).
- Hanley, J.A. & McNeil, B.J. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* **143**, 29–36 (1982).
- Pencina, M.J., D'Agostino, R.B., Sr., D'Agostino, R.B., Jr. & Vasan, R.S. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat. Med.* **27**, 157–172 discussion 207–112 (2008).
- Yachida, S. *et al.* Distant metastasis occurs late during the genetic evolution of pancreatic cancer. *Nature* **467**, 1114–1117 (2010).
- Hingorani, S.R. *et al.* Trp53<sup>R172H</sup> and Kras<sup>G12D</sup> cooperate to promote chromosomal instability and widely metastatic pancreatic ductal adenocarcinoma in mice. *Cancer Cell* **7**, 469–483 (2005).
- Bardeesy, N. *et al.* Both p16<sup>Ink4a</sup> and the p19<sup>Arf</sup>-p53 pathway constrain progression of pancreatic adenocarcinoma in the mouse. *Proc. Natl. Acad. Sci. USA* **103**, 5947–5952 (2006).
- Jackson, E.L. *et al.* The differential effects of mutant p53 alleles on advanced murine lung cancer. *Cancer Res.* **65**, 10280–10288 (2005).
- DuPage, M., Dooley, A.L. & Jacks, T. Conditional mouse lung cancer models using adenoviral or lentiviral delivery of Cre recombinase. *Nat. Protoc.* **4**, 1064–1072 (2009).
- Kirsch, D.G. *et al.* A spatially and temporally restricted mouse model of soft tissue sarcoma. *Nat. Med.* **13**, 992–997 (2007).
- Duell, E.J. *et al.* Pancreatitis and pancreatic cancer risk: a pooled analysis in the International Pancreatic Cancer Case-Control Consortium (PanC4). *Ann. Oncol.* **23**, 2964–2970 (2012).
- Guerra, C. *et al.* Chronic pancreatitis is essential for induction of pancreatic ductal adenocarcinoma by K-Ras oncogenes in adult mice. *Cancer Cell* **11**, 291–302 (2007).
- Gidekel Friedlander, S.Y. *et al.* Context-dependent transformation of adult pancreatic cells by oncogenic K-Ras. *Cancer Cell* **16**, 379–389 (2009).
- Brosnan, J.T. Interorgan amino acid transport and its regulation. *J. Nutr.* **133**, 2068S–2072S (2003).
- Matthews, D.E., Marano, M.A. & Campbell, R.G. Splanchnic bed utilization of leucine and phenylalanine in humans. *Am. J. Physiol.* **264**, E109–E118 (1993).
- Cynober, L.A. Plasma amino acid levels with a note on membrane transport: characteristics, regulation, and metabolic significance. *Nutrition* **18**, 761–766 (2002).
- Acharyya, S. *et al.* Cancer cachexia is regulated by selective targeting of skeletal muscle gene products. *J. Clin. Invest.* **114**, 370–378 (2004).
- Fearon, K.C., Glass, D.J. & Guttridge, D.C. Cancer cachexia: mediators, signaling, and metabolic pathways. *Cell Metab.* **16**, 153–166 (2012).
- Lecker, S.H. *et al.* Multiple types of skeletal muscle atrophy involve a common program of changes in gene expression. *FASEB J.* **18**, 39–51 (2004).
- Lundholm, K., Bylund, A.C., Holm, J. & Schersten, T. Skeletal muscle metabolism in patients with malignant tumor. *Eur. J. Cancer* **12**, 465–473 (1976).
- Heymsfield, S.B. & McManus, C.B. Tissue components of weight loss in cancer patients. A new method of study and preliminary observations. *Cancer* **55**, 238–249 (1985).
- Dewys, W.D. *et al.* Prognostic effect of weight loss prior to chemotherapy in cancer patients. Eastern Cooperative Oncology Group. *Am. J. Med.* **69**, 491–497 (1980).
- Wigmore, S.J., Plester, C.E., Richardson, R.A. & Fearon, K.C. Changes in nutritional status associated with unresectable pancreatic cancer. *Br. J. Cancer* **75**, 106–109 (1997).
- Tsoli, M. & Robertson, G. Cancer cachexia: malignant inflammation, tumorkines, and metabolic mayhem. *Trends Endocrinol. Metab.* **24**, 174–183 (2013).
- Son, J. *et al.* Glutamine supports pancreatic cancer growth through a KRAS-regulated metabolic pathway. *Nature* **496**, 101–105 (2013).
- Commisso, C. *et al.* Macropinocytosis of protein is an amino acid supply route in Ras-transformed cells. *Nature* **497**, 633–637 (2013).
- Harper, A.E., Miller, R.H. & Block, K.P. Branched-chain amino acid metabolism. *Annu. Rev. Nutr.* **4**, 409–454 (1984).
- Shah, S.H., Kraus, W.E. & Newgard, C.B. Metabolomic profiling for the identification of novel biomarkers and mechanisms related to common cardiovascular diseases: form and function. *Circulation* **126**, 1110–1120 (2012).
- Pannala, R., Basu, A., Petersen, G.M. & Chari, S.T. New-onset diabetes: a potential clue to the early diagnosis of pancreatic cancer. *Lancet Oncol.* **10**, 88–95 (2009).

## ONLINE METHODS

**Study population.** Our study population included pancreatic cancer cases and controls from four prospective cohort studies: the Health Professionals Follow-Up Study (HPFS), the Nurses' Health Study (NHS), the Physicians' Health Study I (PHS) and the Women's Health Initiative-Observational Study (WHI-OS). HPFS was initiated in 1986 when 51,529 US men 40–75 years of age working in health professions completed a mailed biennial questionnaire. NHS was established in 1976 when 121,700 female nurses aged 30–55 years completed a mailed biennial questionnaire. PHS is a completed trial initiated in 1982 of aspirin and  $\beta$ -carotene among 22,071 male physicians, aged 40–84 years. After trial completion, study participants were followed as an observational cohort. WHI-OS consists of 93,676 postmenopausal women, aged 50–79 years, enrolled from 1994–1998 at 40 US clinical centers. Participants completed a baseline clinic visit and annual mailed questionnaires. The study was approved by the Human Research Committee at Brigham and Women's Hospital (Boston, MA), and participants provided written informed consent.

We included incident pancreatic adenocarcinoma cases diagnosed after blood collection through 2010 with available plasma and no prior history of cancer. Cases were identified by self-report or follow-up of deaths. Deaths were ascertained from next-of-kin, postal service or National Death Index; this method captures >98% of deaths<sup>43</sup>. Medical records were reviewed by physicians blinded to exposure data to confirm pancreatic cancer diagnoses. Similar to prior studies in these cohorts<sup>4,44–46</sup> and based on a predefined analysis plan, we included only cases diagnosed  $\geq 2$  years after blood collection, as the weight loss and insulin resistance that develop due to pancreatic cancer manifest in the 2 years before diagnosis<sup>42,47</sup>. For each case, we randomly selected two controls, matching on cohort (also matches on sex), year of birth ( $\pm 5$  years), smoking status (never, past, current, missing), fasting status ( $< 8$  h,  $\geq 8$  h), and month/year of blood collection ( $\pm 3$  months in HPFS,  $\pm 3$  months in NHS,  $\pm 6$  months in PHS, and exact matching in WHI). Controls were alive without cancer at the case's diagnosis date and provided a blood sample. Covariate data were obtained from baseline questionnaires in PHS and WHI and questionnaires before blood collection in HPFS and NHS, as described previously<sup>4,45</sup>. In HPFS, NHS, and PHS, cancer stage among cases was directly classified based on medical record review as local disease amenable to surgical resection, locally advanced disease that is unresectable but without distant metastases, or distant metastatic disease. In WHI, medical records were coded using Surveillance Epidemiology End Results summary staging, which classifies tumors as localized, regional, or distant. These stages were then classified in the same manner as in HPFS, NHS and PHS, as local, locally advanced, and metastatic disease, respectively.

The initial data set included 454 cases and 908 controls. Seven controls had insufficient plasma for metabolite profiling. One case and one control were excluded due to missing data for >10% of metabolites.

**Plasma samples.** Blood samples in EDTA tubes were collected from 18,225 men in HPFS from 1993–1995, 14,916 men in PHS from 1982–1984, and 93,676 women in WHI-OS from 1994–1998, and in heparin tubes for 32,826 women in NHS from 1989–1990. Comparisons of participant characteristics in the blood collection cohort and the full cohort are provided for each study in **Supplementary Note**, Table S1. Blood samples in HPFS and NHS were collected by participants, mailed overnight on cold packs, and then spun to collect and store plasma (delayed processing), whereas PHS and WHI participants' whole blood was separately immediately into plasma and stored. An overview of procedures for collection and storage of samples from each cohort is provided below and summarized in **Supplementary Note**, Table S2.

**Health Professionals Follow-up Study.** Upon arrival at the blood lab, vials were centrifuged in order to separate the various component parts. Cryo storage tubes were labeled with the appropriate study member's ID number, and the separated blood components were pipetted into them. This process produced 5 tubes of plasma, 2 tubes of white blood cells, and 1 tube of red blood cells for each cohort member. The tubes were then stored in liquid nitrogen freezers. A bulk tank, holding up to 3,000 gallons of liquid nitrogen, automatically feeds each individual freezer whenever the freezer's sensors indicate that coolant is required.

**Nurses' Health Study.** Blood samples were separated into components (plasma, white blood cells and red blood cells) and pipetted into 8 cryotubes with 5 tubes

of plasma, 2 tubes of white blood cells and 1 tube of red blood cells. Samples were immediately frozen in vapor-phase liquid nitrogen freezers. The NHS Blood Lab stores all biologic samples associated with the Blood Study in-house in a large liquid nitrogen freezer farm. The cryotubes are stored in the vapor phase of liquid-nitrogen freezers; the highest freezer temperature is  $-130$  °C near the top of the freezer, and the lowest temperature is  $-196$  °C at the bottom near the liquid nitrogen. All freezers are alarmed and monitored continuously either by NHS laboratory staff or a central security desk (nights and weekends).

**Physicians' Health Study.** Blood collection kits were sent to all participants with instructions to have blood drawn into the EDTA tubes that were provided. Two tubes were centrifuged for plasma, and a third tube was for whole blood. The specimens were received in the laboratory on chill packs within 24 h of being drawn. Upon receipt, the samples were refrigerated and re-aliquoted into nine 1.2-mL tubes (three whole blood and six plasma), all frozen at  $-82$  °C.

**Women's Health Initiative:** Blood samples were collected on all WHI-OS participants at a baseline clinic visit in the fasting state. Blood samples were maintained at 4 °C for up to one hour until plasma or serum was separated from cells. Centrifuged aliquots were put into  $-70$  °C freezers within two hours of collection. Samples were shipped frozen by overnight delivery to a central facility and kept within  $-70$  °C freezers.

Plasma samples were grouped based on cohort, so that all cases and controls from a single cohort study underwent metabolite profiling as a batch. Sample triplets (pancreatic cancer case, matched control #1, and matched control #2) were distributed randomly within the batch, and the order of the case and two matched controls within each triplet was also randomly designated. Therefore, the case and its two controls were always run in the same batch and were always directly adjacent to each other in the analytic run, thereby limiting variability in platform performance across matched case-control triplets.

For participants from all four cohorts, plasma samples were thawed once to aliquot them from large-volume vials into the smaller volumes needed for shipment to the Broad Institute of the Massachusetts Institute of Technology and Harvard University (Cambridge, MA). The samples were refrozen at the Broad Institute and then thawed a second time to perform metabolite profiling. Therefore, for all cases and controls, plasma samples had been thawed twice at the time of metabolite profiling.

We previously measured hemoglobin A1c (HbA1c) in 389 cases and 757 controls in the laboratory of N. Rifai (Children's Hospital, Boston, MA) using reagents from Roche Diagnostics (Indianapolis, IN). We measured plasma insulin in 386 cases and 743 controls, plasma proinsulin in 388 cases and 746 controls, and plasma C-peptide in 408 cases and 785 controls using reagents from Diagnostic Systems Laboratory (Webster, TX) and Millipore Corporation (Billerica, MA). Randomly inserted samples from quality control (QC) plasma pools had mean intra-assay coefficients of variance (CVs) of 2.0% for HbA1c, 5.4% for insulin, 3.1% for proinsulin, and 4.9% for C-peptide<sup>4</sup>.

**Metabolite profiling.** Profiles of endogenous polar metabolites were obtained using liquid chromatography-tandem mass spectrometry (LC-MS) at the Broad Institute of the Massachusetts Institute of Technology and Harvard University (Cambridge, MA). The LC-MS methods were designed to enable broad measurement of metabolic markers and intermediates, including metabolites from central metabolism and amino acid metabolism, using low plasma sample volumes<sup>48</sup>. LC-MS parameters for targeted analyses, including chromatographic retention times and mass spectrometry multiple reaction monitoring settings (declustering potentials, collision energies, and lens voltages), were determined using over 300 commercially available reference compounds. A subset of 133 polar metabolites were measurable in plasma using a combination of two distinct hydrophilic interaction liquid chromatography (HILIC) methods, one operated under acid mobile phase conditions with positive-ion-mode MS detection and the other under basic elution conditions with negative-ion-mode MS detection.

The acidic HILIC method using positive-ionization-mode MS analyses was similar to the method described by Wang *et al.*<sup>11</sup>. Briefly, the LC-MS system consisted of a 4000 QTRAP triple quadrupole mass spectrometer (AB SCIEX) coupled to an 1100 Series pump (Agilent) and an HTS PAL autosampler (Leap Technologies). Plasma samples (10  $\mu$ L) were extracted using nine volumes of 74.9:24.9:0.2 (v/v/v) acetonitrile/methanol/formic acid containing stable

isotope-labeled internal standards (valine-d8, Isotec; and phenylalanine-d8, Cambridge Isotope Laboratories). The samples were centrifuged (10 min, 9,000g, 4 °C), and the supernatants (10  $\mu$ L) were injected onto an Atlantis HILIC column (150  $\times$  2.1 mm, 3  $\mu$ m particle size; Waters Inc.). The column was eluted isocratically at a flow rate of 250  $\mu$ L/min with 5% mobile phase A (10 mM ammonium formate and 0.1% formic acid in water) for 1 min followed by a linear gradient to 40% mobile phase B (acetonitrile with 0.1% formic acid) over 10 min. The ion spray voltage was 4.5 kV and the source temperature was 450 °C.

A second method using basic HILIC separation and negative ionization mode MS detection was established on an LC-MS system consisting of an ACQUITY UPLC (Waters Inc.) coupled to a 5500 QTRAP triple quadrupole mass spectrometer (AB SCIEX). Plasma samples (30  $\mu$ L) were extracted using 120  $\mu$ L of 80% methanol (VWR) containing the internal standards inosine-<sup>15</sup>N<sub>4</sub>, thymine-d<sub>4</sub>, and glycocholate-d<sub>4</sub> (Cambridge Isotope Laboratories). The samples were centrifuged (10 min, 9,000g, 4 °C), and the supernatants were injected directly onto a Luna NH<sub>2</sub> column (150  $\times$  2.0 mm, 5  $\mu$ m particle size; Phenomenex) that was eluted at a flow rate of 400  $\mu$ L/min with initial conditions of 10% mobile phase A (20 mM ammonium acetate and 20 mM ammonium hydroxide (Sigma-Aldrich) in water (VWR)) and 90% mobile phase B (10 mM ammonium hydroxide in 75:25 v/v acetonitrile/methanol (VWR)) followed by a 10-min linear gradient to 100% mobile phase A. The ion spray voltage was -4.5 kV and the source temperature was 500 °C.

Raw data were processed using MultiQuant 1.2 software (AB SCIEX) for automated LC-MS peak integration. All chromatographic peaks were manually reviewed for quality of integration and compared against a known standard for each metabolite to confirm compound identities. Internal standard peak areas were monitored for quality control, to assess system performance over time, and to identify any outlier samples requiring re-analysis. A pooled plasma reference sample was also analyzed after sets of 20 study samples as an additional quality control measure of analytical performance and to serve as reference for scaling raw LC-MS peak areas across sample batches. Metabolites with a signal-to-noise ratio <10 were considered unquantifiable. Metabolite signals were analyzed in relation to pancreatic cancer risk as LC-MS peak areas, which are proportional to metabolite concentration and appropriate for metabolite clustering and correlative analyses.

Of the 133 metabolites measured (Supplementary Note, Fig. S1), 83 were included in the analyses of our nested pancreatic cancer case-control population (Supplementary Note, Table S3). In pilot work<sup>49</sup>, we determined that 32 metabolites had poor reproducibility in samples with delayed processing, so these metabolites were excluded as they could not be reliably measured in two of the participating cohorts. In the current study, three heparin plasma pools (57 total QC samples) and three EDTA plasma pools (128 total QC samples) were randomly interspersed among participant samples as blinded QC samples. We calculated mean CVs for each metabolite across QC plasma pools and set an a priori threshold of >25% for satisfactory reproducibility. Based on this criterion, 13 metabolites with mean CV>25% were excluded from our analyses. Five metabolites had undetectable levels for >10% of participants and were also excluded. We evaluated plasma from ten volunteers with plasma collected simultaneously in heparin and EDTA tubes. For the branched chain amino acids, Spearman correlation coefficients between Heparin and EDTA samples were 0.85 for isoleucine, 0.88 for leucine, and 0.95 for valine.

For metabolites meeting the threshold for statistical significance after multiple-hypothesis correction (isoleucine, leucine and valine), LC-MS peak areas were converted to absolute concentrations using stable isotope-labeled standards. Briefly, external calibration curves of MS response were determined using solutions of isotope-labeled <sup>13</sup>C<sub>6</sub>, <sup>15</sup>N-leucine, <sup>13</sup>C<sub>6</sub>, <sup>15</sup>N-isoleucine (Cambridge Isotope Laboratories), and d<sub>8</sub>-valine (Isotec). A 1  $\mu$ g/ $\mu$ L solution of each standard was prepared in water. 20  $\mu$ L of each stock solution were added to 180  $\mu$ L of reference pooled plasma, and the resulting solution was then serially diluted using pooled plasma to generate a calibration curve. For multiple reaction monitoring MS analyses, the bond cleavage products and collision energy (CE) and declustering potential (DP) settings were the same as those used for the endogenous metabolites: natural leucine 132/86, CE = 18 and DP = 50; <sup>13</sup>C<sub>6</sub>, <sup>15</sup>N-leucine 134/87, CE = 18 and DP = 50; natural isoleucine 132/86, CE = 18 and DP = 50; <sup>13</sup>C<sub>6</sub>, <sup>15</sup>N-isoleucine 139/92, CE = 18 and DP = 50; natural valine 118/72, CE = 18 and DP = 25, and d<sub>8</sub>-valine 126/80, CE = 18 and DP = 25.

Three separate plasma samples were prepared at each concentration and were analyzed using the acidic HILIC LC-MS method described above. The median concentrations of endogenous isoleucine, leucine and valine in the reference pooled plasma were calculated, and the concentration of each metabolite in each study sample was determined from the response ratio relative to the nearest reference pooled plasma sample in the analysis queue.

**Statistical analyses for human studies.** To compare baseline characteristics, we used conditional logistic regression conditioned on the matching factors and including the covariate of interest. Partial Spearman correlation coefficients were calculated for metabolites and covariates. Metabolites were log-transformed to improve normality and included as continuous variables in conditional logistic regression models conditioned on matching factors and adjusted for age at blood draw (years, continuous), fasting time (< 4 h, 4–8 h, 8–12 h,  $\geq$ 12 h, missing) and race or ethnicity (white, black, other, missing). Using a conservative Bonferroni correction for multiple-hypothesis testing<sup>50</sup>, metabolites with  $P < 0.0006$  (0.05/83) were considered statistically significant.

To provide estimates of effect magnitude, significant metabolites were examined in conditional logistic regression models after categorization into quintiles based on log-transformed metabolite levels in controls. Separate quintiles were generated for fasting ( $\geq$ 8 h since last meal) and nonfasting (<8 h since last meal) participants, given the possible effect of fasting time on metabolite levels. Quintiles were generated from the population of selected controls, which may not exactly reflect the characteristics of the full cohort population. Odds ratios (ORs) and 95% confidence intervals (CIs) were also calculated per s.d. change in log-transformed metabolite levels. To control for possible confounding, we evaluated regression models adjusted for body-mass index (BMI), physical activity, history of diabetes mellitus, HbA1c, plasma insulin, plasma proinsulin, and plasma C-peptide. We also evaluated regression models after excluding subjects with diabetes by self-report or HbA1c  $\geq$ 6.5% at blood collection (prevalent diabetes). We further evaluated models that excluded subjects who developed diabetes after blood collection but >2 years before cancer diagnosis (incident diabetes not thought to be recent onset from pancreatic cancer)<sup>42,47</sup>.

Metabolite values were considered missing when an LC-MS peak was below the limit of detection. In the primary analysis, any case or control with missing data for a metabolite was excluded from the analysis of that metabolite. However, we also conducted sensitivity analyses, in which participants with missing values were assigned the lower limit of detection or half of the lower limit of detection, and our results were unchanged.

We assessed heterogeneity of metabolite associations with pancreatic cancer risk across cohorts using Cochran's  $Q$ -statistic<sup>51</sup>. We examined associations in predefined subgroups by sex, smoking status, BMI, and fasting status. Statistical interactions were assessed by entering into models the main effect terms and cross-product terms of metabolites and stratification variables, evaluating likelihood ratio tests. We also examined associations by time between blood collection and the case's cancer diagnosis. In these time-based analyses, one stratum included 40 pancreatic cancer cases with blood collected within 2 years of diagnosis and their matched controls. These cases and controls were not part of the primary analysis population, but were included in the stratified analyses by time to more fully delineate the association of metabolites with pancreatic development by time before diagnosis. Associations were also examined for circulating BCAAs with previously explored risk factors for pancreatic cancer in our cohorts (Supplementary Note, Table S4) and with cancer stage at diagnosis (Supplementary Note, Table S5).

Since association of a marker with disease does not indicate the suitability of the marker to serve as a screening test for the disease, we examined two approaches to quantify the value of metabolites in a multifactor risk discrimination tool for pancreatic cancer. Discrimination quantifies the ability of one or more disease markers to separate cases (individuals with the disease) from controls (individuals without the disease). We investigated the discrimination of risk models for predicting pancreatic cancer diagnosis in the 10 years after measurement of circulating BCAAs, i.e., the 10-year risk of pancreatic cancer.

In the first approach, we investigated receiver-operating-characteristic (ROC) curve analysis and calculated of the area under the ROC curve (AUC), also known as the concordance (C) statistic<sup>16,52</sup>. The base model included age at blood collection (continuous), cohort (HPFS, NHS, PHS, WHI; which also

accounts for sex), race/ethnicity (white, black, other/missing), smoking status (never, past, current, missing) and fasting time (<4 h, 4–8 h, 8–12 h, ≥12 h, missing). Three subsequent models mirrored the base model but additionally included (1) body-mass index, physical activity, and history of diabetes, (2) circulating BCAAs, or (3) body-mass index, physical activity, history of diabetes, and circulating BCAAs. Each point on the ROC curve shows the effect of a rule for turning a risk estimate into a prediction of the development of an event. The y axis of the ROC curve is the true positive rate or sensitivity (i.e., the proportion of individuals with pancreatic cancer who were correctly predicted to have the disease). The x axis shows the false positive rate, which is the complement of specificity (i.e., the proportion of individuals without pancreatic cancer who were incorrectly predicted to have pancreatic cancer). The area under the ROC curve, the AUC, measures how well the model discriminates between case subjects and control subjects. An ROC curve that corresponds to a random classification of case subjects and control subjects is a straight line with an AUC of 50%. An ROC curve that corresponds to perfect classification has an AUC of 100%. The improvement in AUC for a model containing a new marker is defined as the difference in AUCs calculated using a model with and without the new marker of interest<sup>53</sup>.

For context, the Breast Cancer Risk Assessment Tool, commonly referred to as the Gail model<sup>54,55</sup>, estimates a woman's risk for breast cancer using clinically available information including current age, age at menarche, age of first live birth, number of first-degree relatives with breast cancer, number of previous breast biopsies, breast biopsies that show atypical hyperplasia, and race/ethnicity. The Gail model is used to counsel women on appropriate screening tests for breast cancer<sup>56</sup>, for determining whether tamoxifen will be useful as a chemopreventative agent<sup>57</sup>, and for determining sample size calculations in randomized clinical trials of prevention strategies<sup>58</sup>. Several studies have evaluated the discrimination of the Gail model using ROC curve analysis and calculated the AUC to be 0.58 to 0.63 (refs. 52,59–61). Follow-up studies have described an AUC of 0.62 to 0.66 when breast density is added as an additional predictor in the original Gail model<sup>59,60,62</sup>.

Although ROC curves are commonly used, they have a number of limitations and may underestimate the ability of a new marker to contribute to risk prediction when added to previously defined predictors<sup>63–66</sup>. Another approach to evaluating model discrimination is to evaluate the ability of a new marker to shift an individual's risk up or down between predefined risk categories. This is known as the prediction increment of a marker and has been codified in an approach known as net reclassification improvement (NRI)<sup>17</sup>. The NRI (sometimes referred to as the net reclassification index) constructs reclassification tables separately for participants with and without events and quantifies the correct movement between categories of risk, namely, to higher risk categories for participants with events and to lower risk categories for those without events. Furthermore, incorrect movement in categories of risk (downwards for events and upwards for non-events) reduces the net correct reclassification of individuals within the study population.

The NRI calculation is represented by the following formula:

$$\text{NRI} = [P(\text{up}|D = 1) - P(\text{down}|D = 1)] + [P(\text{up}|D = 0) - P(\text{down}|D = 0)]$$

Upward movement (up) is defined as a change into a higher risk category based on the new model and downward movement (down) as a change into a lower risk category based on the new model, where *P* indicates probability and *D* denotes the event indicator (1, event; 0, non-event).

Using the NRI, we evaluated the ability of the prediction model including circulating BCAAs to appropriately reclassify individuals into risk groups compared to the base model. The base model was calculated using conditional logistic regression conditioned on matching factors and adjusted for race/ethnicity, body-mass index, physical activity and history of diabetes. The subsequent model included the covariates in the base model with the addition of circulating metabolites. As in prior studies<sup>67–69</sup>, we defined the high-risk group as those individuals with risk for pancreatic cancer at least twofold greater than an individual with average risk.

For context, the Emerging Risk Factors Collaboration<sup>70</sup> has examined the integration of novel risk factors into risk prediction models for cardiovascular disease. In these studies, additional potential risk predictors were added to a

model of known risk predictors for cardiovascular disease, including age, sex, smoking status, blood pressure, history of diabetes, and cholesterol. The net reclassification improvement was then calculated for three 10-year risk categories for cardiovascular disease. C-reactive protein (CRP) is a marker of systemic inflammation, and elevated CRP has been associated with an increased risk for cardiovascular events in numerous studies<sup>71–73</sup>. Circulating CRP is currently used to inform decisions in the clinic regarding screening and risk reduction strategies<sup>74,75</sup> and to design clinical trials testing novel treatments to reduce cardiovascular events<sup>76,77</sup>. In an analysis of nearly 250,000 individuals<sup>78</sup>, the addition of CRP to known cardiovascular disease risk factors was associated with a statistically significant improvement in the area under the ROC curve and a NRI of 1.52% for 10-year risk of cardiovascular disease. In contrast, additional analyses demonstrated a <1% improvement in the NRI for body-mass index, waist circumference, waist-to-hip ratio, plasma fibrinogen, and circulating apolipoproteins<sup>78–80</sup>, such that the clinical utility of these additional predictors remains unclear<sup>75</sup>.

All analyses were performed with SAS 9.2 statistical package. All *P* values were two-sided.

**Experimental mice.** All studies were approved by the MIT Committee on Animal Care (IACUC). All experimental groups were assigned based on genotype. All animals were numbered and experiments conducted blinded. After data collection, genotypes were revealed and animals assigned to groups for analysis. The experiments were not randomized.

**KPC.** Experimental KPC mice were male mice on a mixed background, heterozygous for the conditional lox-stop-lox *Kras*<sup>G12D</sup> allele, heterozygous for the conditional lox-stop-lox *Trp53*<sup>R172H</sup> allele and expressing Cre recombinase under control of the *Pdx1* promoter (*Tg(Ipfl-cre)I<sup>Tuv</sup>*)<sup>19</sup>. Littermate controls lacked either the LSL-*Kras*<sup>G12D</sup> allele, the Cre allele or both. Control mice were killed at the same time as their tumor-bearing littermates.

**KP<sup>-/-</sup>C.** Experimental KP<sup>-/-</sup>C mice were male mice on a mixed background, heterozygous for the conditional lox-stop-lox *Kras*<sup>G12D</sup> allele, homozygous for loxP sites flanking exons 2–10 of *Trp53* and expressing Cre recombinase under control of the *Pdx1* promoter (*Tg(Ipfl-cre)I<sup>Tuv</sup>*)<sup>20</sup>. Littermate control mice lacked either the Cre-recombinase allele, LSL-*Kras*<sup>G12D</sup> allele or both (controls were non-tumor-bearing mice of all genotypes). Inbred C57BL/6J male mice containing the same alleles were also examined where indicated, and cancer cell lines derived from these mice (established in culture from tumors prior to the described implantation studies) were used for syngenic implantation studies.

**Non-small-cell lung cancer.** Six-month-old male mice on a pure 129 background, heterozygous for the conditional lox-stop-lox *Kras*<sup>G12D</sup> allele and homozygous for loxP sites flanking exons 2–10 of *Trp53*, were administered  $2.5 \times 10^7$  PFU of Cre-expressing adenovirus intratracheally as previously described<sup>21,22</sup>. High-titer adenovirus was obtained from the Gene Transfer Vector Core (University of Iowa).

**Hindlimb sarcoma.** Four-week-old male mice on a mixed background, heterozygous for the conditional lox-stop-lox *Kras*<sup>G12D</sup> allele and homozygous for loxP sites flanking exons 2–10 of *Trp53* were administered  $2.5 \times 10^8$  PFU of Cre-expressing adenovirus intramuscularly as previously described<sup>23</sup>. High-titer adenovirus was obtained from the Gene Transfer Vector Core (University of Iowa).

**Implantation, pancreatitis and BCAA diet studies.** Male C57BL/6J mice aged 4–6 weeks at the start of the study were used for these experiments.

**Diets.** Standard chow diet was RMH 3000 (Prolab). For amino acid-defined diets, 1× BCAA (TD.110839) and 2× BCAA (TD.110843) were designed in consultation with and subsequently obtained from Harlan Teklad. 20% <sup>13</sup>C-leucine- and 20% <sup>13</sup>C-valine-labeled diets were based on diet TD.110839 and produced by Cambridge Isotopes and Harlan Teklad.

**Plasma for metabolomics.** Plasma was collected for each experiment at the time points indicated. Mice were anesthetized under 2% isoflurane-oxygen mixture and retro-orbitally bled approximately 4.5 h after the onset of the light cycle. Blood was immediately placed in EDTA-pretreated tubes and centrifuged to separate plasma. Plasma was aliquoted and frozen at –80 °C for further analysis.

Fasting blood samples were harvested in the same manner first thing in the morning after a 16-hour overnight fast.

**Food consumption.** Mice were housed individually for 48 h, and remaining food pellets weighed at 0, 24 and 48 h. A two-day average was then calculated for each mouse. Body weight was determined on the second day. To assess consumption of BCAA defined diets, mice were housed individually and fed diets for 5 d. Food was weighed after 2 d of feeding and again on day 5, and the average consumption per 24 h over the 72-h period was calculated.

**Blood glucose, plasma insulin, glucose tolerance test and insulin tolerance test.** Four-week-old  $KP^{-/-}$  mice were fasted overnight and blood glucose measured using a One Touch Ultra handheld glucometer. 25  $\mu$ L of plasma from the same mice was harvested in heparinized tubes, aliquoted, and frozen at  $-80^{\circ}\text{C}$  for further analysis. Plasma insulin levels were determined using an ultrasensitive mouse insulin ELISA kit (Crystal Chem, #90080). After measuring fasting parameters, a glucose tolerance test was performed in accordance with published protocols<sup>81</sup>. Briefly, conscious mice received an intraperitoneal injection of 2 g/kg glucose at time 0. Blood glucose was subsequently measured at 15, 30, 60, 90 and 120 min post-injection as described above. For insulin tolerance test, 4-week-old  $KP^{-/-}$  mice were fasted for 6 h during daytime hours. Following initial blood glucose measurement, conscious mice received an intraperitoneal injection of 0.75 IU/kg recombinant human insulin (Novolin, Novo Nordisk). Blood glucose was subsequently measured at 15, 30, 60 and 90 min post-injection as described above.

**$KP^{-/-}$  cell lines and implantation studies.** End-stage tumors were dissected from C57BL/6J  $KP^{-/-}$  mice and mechanically chopped before trypsin disaggregation, with tumor cells then propagated for three to five passages in DMEM with 10% FBS, 4 mM glutamine and penicillin/streptomycin to obtain enough cells for implantation. Cell lines were negative for mycoplasma. For subcutaneous implantation studies, recipient mice were anesthetized with inhaled 2% isoflurane-oxygen mixture, low passage cell lines (passage 5 for each line) were resuspended at  $2.5 \times 10^5$  cells per 100  $\mu$ L in sterile PBS, and 100  $\mu$ L of either PBS (control) or cell suspension was injected in the flank of syngeneic mice. For orthotopic implantation studies, recipient mice were anesthetized with inhaled 2% isoflurane-oxygen mixture, a vertical incision made in the abdomen at the left mid-calvicular line, the spleen mobilized, and 25  $\mu$ L of either PBS or PBS containing  $1.0 \times 10^5$  cells (passage 3 for each line) was injected into the tail of the pancreas.

**Caerulein-induced chronic pancreatitis.** Mice were treated with either USP-grade saline or 5  $\mu$ g caerulein (Sigma) via intraperitoneal injection daily, 5 d per week for 10 weeks as previously described<sup>25</sup>. Blood was obtained and the mice killed on the final day. Tissues were fixed in 10% formalin for subsequent histological analysis.

**Plasma markers of pancreatitis.** Plasma amylase and lipase measurements were performed by IDEXX BioResearch Laboratory (North Grafton, MA).

**BCAA diet studies.** Mice were fed either 1 $\times$  or 2 $\times$  BCAA diets for 10 weeks. Blood was obtained and mice killed on the final day of the experiment. Tissues were fixed in 10% formalin for subsequent histological analysis.

**Studies to determine source of BCAA elevations. Acute uptake and disposal.** Following a 16-h overnight fast, mice were fed 20%  $^{13}\text{C}$ -leucine and valine containing diet for 2 h before removal of food, and food consumption during this period quantified as described above. At the time points indicated by the red arrowheads in **Figure 3a**, 10–25  $\mu$ L of plasma was harvested from the tail vein of conscious mice in a heparinized tube and centrifuged to separate plasma. Plasma was aliquoted and frozen at  $-80^{\circ}\text{C}$  for subsequent GC-MS analysis. Total ion counts from GC-MS analysis of leucine and valine were then normalized to norvaline internal standard and multiplied by fractional labeling to determine the amount of label present. This number for each animal was then normalized to that animal's food intake to control for interanimal variation in labeled food consumption.

**Long-term pool contribution of BCAA to plasma.** Mice were exposed to 20%  $^{13}\text{C}$ -leucine- and  $^{13}\text{C}$ -valine-labeled diets from 7 d of age to 24 d of age followed by 3 d of unlabeled diet (according to the protocol depicted in **Fig. 3b**). Two cohorts of mice were used in this study. One cohort of mice was killed on day 27 in the fed state, and a second cohort of mice was killed on day 28 after a 16-h overnight fast (the points indicated by the red arrowheads in **Fig. 3b**). At time of killing, anesthetized mice were terminally bled and tissues harvested within 5 min, snap frozen in liquid nitrogen using Biosqueezer (BioSpec Products), and stored at  $-80^{\circ}\text{C}$  for subsequent GC-MS analysis. Plasma was aliquoted and frozen at  $-80^{\circ}\text{C}$  for subsequent GC-MS analysis.

Total contributions from short and long-term pools were calculated according to the following equations:

$$\text{Long Term Pool} = \left\{ \frac{\text{Fed \% Labeled}}{\text{Fasted \% Labeled}} \right\} \cdot \text{Relative Fed Pool Size}$$

$$\text{Short Term Pool} = \text{Relative Fed Pool Size} \square \text{Long Term Pool}$$

Raw data are summarized in **Supplementary Note**, Table S6.

**Tissue and body weights.** For measurement of tissue weights, mice were weighed before killing, then gastrocnemius, tibialis anterior, soleus and heart were subsequently dissected and weighed. Muscle weights for each individual mouse were normalized to the body weight of that mouse.

**LC-MS plasma amino acid measurements.** Plasma amino acids were measured by LC-MS at the Koch Institute of the Massachusetts Institute of Technology (Cambridge, MA) using similar methods used for assessment of metabolites in human plasma. Raw data were analyzed as peak area tops using the open-access MAVEN software tool<sup>82</sup>.

**GC-MS assessment of  $^{13}\text{C}$ -leucine and  $^{13}\text{C}$ -valine labeling.** Plasma polar metabolites were extracted in ice-cold 4:1 methanol/water with norvaline internal standard (5  $\mu$ L plasma in 200  $\mu$ L extraction solution). Extracts were clarified by centrifugation and the supernatant evaporated under nitrogen and frozen at  $-80^{\circ}\text{C}$  for subsequent derivitization. Dried polar metabolites were dissolved in 20  $\mu$ L of 2% methoxyamine hydrochloride in pyridine (Thermo) and held at  $37^{\circ}\text{C}$  for 1.5 h. After dissolution and reaction, tert-butyldimethylsilyl derivatization was initiated by adding 25  $\mu$ L *N*-methyl-*N*-(tert-butyldimethylsilyl)tri fluoroacetamide + 1% tert-butyldimethylchlorosilane (Sigma) and incubating at  $37^{\circ}\text{C}$  for 1 h. The acid hydrolysis protocol was adapted from Antoniewicz *et al.*<sup>83</sup>. Briefly, acid hydrolysis of tissue proteins was performed on snap-frozen tissues by boiling 1–5 mg tissue in 1 mL 18% hydrochloric acid overnight at  $100^{\circ}\text{C}$ . 50  $\mu$ L supernatant was evaporated under nitrogen and frozen at  $-80^{\circ}\text{C}$  for subsequent derivitization. Dried hydrolysates were re-dissolved in pyridine (10  $\mu$ L/1 mg tissue) before tert-butyldimethylsilyl derivatization, which was initiated by adding *N*-methyl-*N*-(tert-butyldimethylsilyl)trifluoroacetamide + 1% tert-butyldimethylchlorosilane (12.5  $\mu$ L/1 mg tissue, Sigma) and incubating at  $37^{\circ}\text{C}$  for 1 h.

GC-MS analysis was performed using an Agilent 7890 GC equipped with a 30-m DB-35MS capillary column connected to an Agilent 5975B MS operating under electron impact ionization at 70 eV. One microliter of sample was injected in splitless mode at  $270^{\circ}\text{C}$ , using helium as the carrier gas at a flow rate of 1 mL  $\text{min}^{-1}$ . For measurement of amino acids, the GC oven temperature was held at  $100^{\circ}\text{C}$  for 3 min and increased to  $300^{\circ}\text{C}$  at  $3.5^{\circ}\text{C min}^{-1}$ . The MS source and quadrupole were held at  $230^{\circ}\text{C}$  and  $150^{\circ}\text{C}$ , respectively, and the detector was run in scanning mode, recording ion abundance in the range of 100–605 *m/z*. MIDs were determined by integrating the appropriate ion fragments<sup>83</sup> listed in **Supplementary Note**, Table S7 and corrected for natural isotope abundance using an algorithm adapted from Fernandez *et al.*<sup>84</sup>.

**Statistical analyses for mouse studies.** Appropriate statistical tests were performed where required. Two-sided unpaired Student's *t*-tests were performed for all statistical analyses unless otherwise specified using Microsoft Excel for Mac:2011 (Microsoft) or GraphPad Prism 6 (GraphPad Software). Two-sided

repeated-measures analysis of variance was performed to compare mean plasma glucose levels in the glucose tolerance test and insulin tolerance tests<sup>85</sup>, using SAS 9.2 statistical package. No statistical method was used to predetermine sample size.

43. Rich-Edwards, J.W., Corsano, K.A. & Stampfer, M.J. Test of the National Death Index and Equifax Nationwide Death Search. *Am. J. Epidemiol.* **140**, 1016–1019 (1994).
44. Michaud, D.S. *et al.* Prediagnostic plasma C-peptide and pancreatic cancer risk in men and women. *Cancer Epidemiol. Biomarkers Prev.* **16**, 2101–2109 (2007).
45. Wolpin, B.M. *et al.* Plasma 25-hydroxyvitamin D and risk of pancreatic cancer. *Cancer Epidemiol. Biomarkers Prev.* **21**, 82–91 (2012).
46. Bao, Y. *et al.* A prospective study of plasma adiponectin and pancreatic cancer risk in five US cohorts. *J. Natl. Cancer Inst.* **105**, 95–103 (2013).
47. Pannala, R. *et al.* Temporal association of changes in fasting blood glucose and body mass index with diagnosis of pancreatic cancer. *Am. J. Gastroenterol.* **104**, 2318–2325 (2009).
48. Roberts, L.D., Souza, A.L., Gerszten, R.E. & Clish, C.B. Targeted metabolomics. *Curr. Protoc. Mol. Biol.* **98**, 30.2 (2012).
49. Townsend, M.K. *et al.* Reproducibility of metabolomic profiles among men and women in 2 large cohort studies. *Clin. Chem.* **59**, 1657–1667 (2013).
50. Bland, J.M. & Altman, D.G. Multiple significance tests: the Bonferroni method. *BMJ* **310**, 170 (1995).
51. Cochran, W.G. The combination of estimates from different experiments. *Biometrics* **10**, 101–129 (1954).
52. Wacholder, S. *et al.* Performance of common genetic variants in breast-cancer risk models. *N. Engl. J. Med.* **362**, 986–993 (2010).
53. DeLong, E.R., DeLong, D.M. & Clarke-Pearson, D.L. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* **44**, 837–845 (1988).
54. Gail, M.H. *et al.* Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. *J. Natl. Cancer Inst.* **81**, 1879–1886 (1989).
55. Gail, M.H. & Benichou, J. Validation studies on a model for breast cancer risk. *J. Natl. Cancer Inst.* **86**, 573–575 (1994).
56. Saslow, D. *et al.* American Cancer Society guidelines for breast screening with MRI as an adjunct to mammography. *CA Cancer J. Clin.* **57**, 75–89 (2007).
57. Gail, M.H. *et al.* Weighing the risks and benefits of tamoxifen treatment for preventing breast cancer. *J. Natl. Cancer Inst.* **91**, 1829–1846 (1999).
58. Costantino, J.P. *et al.* Validation studies for models projecting the risk of invasive and total breast cancer incidence. *J. Natl. Cancer Inst.* **91**, 1541–1548 (1999).
59. Tice, J.A. *et al.* Using clinical factors and mammographic breast density to estimate breast cancer risk: development and validation of a new predictive model. *Ann. Intern. Med.* **148**, 337–347 (2008).
60. Barlow, W.E. *et al.* Prospective breast cancer risk prediction model for women undergoing screening mammography. *J. Natl. Cancer Inst.* **98**, 1204–1214 (2006).
61. Rockhill, B., Spiegelman, D., Byrne, C., Hunter, D.J. & Colditz, G.A. Validation of the Gail *et al.* model of breast cancer risk prediction and implications for chemoprevention. *J. Natl. Cancer Inst.* **93**, 358–366 (2001).
62. Chen, J. *et al.* Projecting absolute invasive breast cancer risk in white women with a model that includes mammographic density. *J. Natl. Cancer Inst.* **98**, 1215–1226 (2006).
63. Cook, N.R. Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation* **115**, 928–935 (2007).
64. Pepe, M.S., Janes, H., Longton, G., Leisenring, W. & Newcomb, P. Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. *Am. J. Epidemiol.* **159**, 882–890 (2004).
65. Cook, N.R., Buring, J.E. & Ridker, P.M. The effect of including C-reactive protein in cardiovascular risk prediction models for women. *Ann. Intern. Med.* **145**, 21–29 (2006).
66. Ridker, P.M., Buring, J.E., Rifai, N. & Cook, N.R. Development and validation of improved algorithms for the assessment of global cardiovascular risk in women: the Reynolds Risk Score. *J. Am. Med. Assoc.* **297**, 611–619 (2007).
67. Roberts, N.J. *et al.* The predictive capacity of personal genome sequencing. *Sci. Transl. Med.* **4**, 133ra158 (2012).
68. van Ravesteyn, N.T. *et al.* Tipping the balance of benefits and harms to favor screening mammography starting at age 40 years: a comparative modeling study of risk. *Ann. Intern. Med.* **156**, 609–617 (2012).
69. Chatterjee, N. *et al.* Projecting the performance of risk prediction based on polygenic analyses of genome-wide association studies. *Nat. Genet.* **45**, 400–405 (2013).
70. Danesh, J. *et al.* The Emerging Risk Factors Collaboration. analysis of individual data on lipid, inflammatory and other markers in over 1.1 million participants in 104 prospective studies of cardiovascular diseases. *Eur. J. Epidemiol.* **22**, 839–869 (2007).
71. Ridker, P.M., Hennekens, C.H., Buring, J.E. & Rifai, N. C-reactive protein and other markers of inflammation in the prediction of cardiovascular disease in women. *N. Engl. J. Med.* **342**, 836–843 (2000).
72. Ridker, P.M. Clinical application of C-reactive protein for cardiovascular disease detection and prevention. *Circulation* **107**, 363–369 (2003).
73. Pradhan, A.D. *et al.* Inflammatory biomarkers, hormone replacement therapy, and incident coronary heart disease: prospective analysis from the Women's Health Initiative observational study. *J. Am. Med. Assoc.* **288**, 980–987 (2002).
74. Ridker, P.M. *et al.* C-reactive protein levels and outcomes after statin therapy. *N. Engl. J. Med.* **352**, 20–28 (2005).
75. Greenland, P. *et al.* 2010 ACCF/AHA guideline for assessment of cardiovascular risk in asymptomatic adults: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. *J. Am. Coll. Cardiol.* **56**, e50–e103 (2010).
76. Ridker, P.M. *et al.* Rosuvastatin to prevent vascular events in men and women with elevated C-reactive protein. *N. Engl. J. Med.* **359**, 2195–2207 (2008).
77. Ridker, P.M. *et al.* Reduction in C-reactive protein and LDL cholesterol and cardiovascular event rates after initiation of rosuvastatin: a prospective study of the JUPITER trial. *Lancet* **373**, 1175–1182 (2009).
78. Kaptoge, S. *et al.* C-reactive protein, fibrinogen, and cardiovascular disease prediction. *N. Engl. J. Med.* **367**, 1310–1320 (2012).
79. Wormser, D. *et al.* Separate and combined associations of body-mass index and abdominal adiposity with cardiovascular disease: collaborative analysis of 58 prospective studies. *Lancet* **377**, 1085–1095 (2011).
80. Di Angelantonio, E. *et al.* Lipid-related markers and cardiovascular disease prediction. *J. Am. Med. Assoc.* **307**, 2499–2506 (2012).
81. Ayala, J.E. *et al.* Standard operating procedures for describing and performing metabolic tests of glucose homeostasis in mice. *Dis. Model. Mech.* **3**, 525–534 (2010).
82. Clasquin, M.F., Melamud, E. & Rabinowitz, J.D. LC-MS data processing with MAVEN: a metabolomic analysis and visualization engine. *Curr. Protoc. Bioinformatics* **37**, 14.11 (2012).
83. Antoniewicz, M.R. *et al.* Metabolic flux analysis in a nonstationary system: fed-batch fermentation of a high yielding strain of *E. coli* producing 1,3-propanediol. *Metab. Eng.* **9**, 277–292 (2007).
84. Fernandez, C.A., Des Rosiers, C., Previs, S.F., David, F. & Brunengraber, H. Correction of <sup>13</sup>C mass isotopomer distributions for natural stable isotope abundance. *J. Mass Spectrom.* **31**, 255–262 (1996).
85. Andrikopoulos, S., Blair, A.R., Deluca, N., Fam, B.C. & Proietto, J. Evaluating the glucose tolerance test in mice. *Am. J. Physiol. Endocrinol. Metab.* **295**, E1323–E1332 (2008).

# insulin-like Growth Factor Pathway Genetic Polymorphisms, circulating iGF1 and iGFBP3, and Prostate cancer Survival

Yin Cao, Sara Lindström, Fredrick Schumacher, Victoria L. Stevens, Demetrius Albanes, Sonja I. Berndt, Heiner Boeing, H. Bas Bueno-de-Mesquita, Federico Canzian, Saioa Chamosa, Stephen J. Chanock, W. Ryan Diver, Susan M. Gapstur, J. Michael Gaziano, Edward L. Giovannucci, Christopher A. Haiman, Brian Henderson, Mattias Johansson, Loïc Le Marchand, Domenico Palli, Bernard Rosner, Afshan Siddiq, Meir Stampfer, Daniel O. Stram, Rulla Tamimi, Ruth C. Travis, Dimitrios Trichopoulos, Walter C. Willett, Meredith Yeager, Peter Kraft, Ann W. Hsing, Michael Pollak, Xihong Lin, Jing Ma

Manuscript received November 18, 2013; revised March 3, 2014; accepted March 4, 2014.

**Correspondence to:** Jing Ma, MD, PhD, Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, 181 Longwood Ave, Boston, MA 02115 (e-mail: [jing.ma@channing.harvard.edu](mailto:jing.ma@channing.harvard.edu)) and Yin Cao, MPH, ScD, Department of Nutrition, Harvard School of Public Health, 665 Huntington Ave, Boston, MA 02115 (e-mail: [yic576@mail.harvard.edu](mailto:yic576@mail.harvard.edu)).

**Background** The insulin-like growth factor (IGF) signaling pathway has been implicated in prostate cancer (PCa) initiation, but its role in progression remains unknown.

**Methods** Among 5887 PCa patients (704 PCa deaths) of European ancestry from seven cohorts in the National Cancer Institute Breast and Prostate Cancer Cohort Consortium, we conducted Cox kernel machine pathway analysis to evaluate whether 530 tagging single nucleotide polymorphisms (SNPs) in 26 IGF pathway-related genes were collectively associated with PCa mortality. We also conducted SNP-specific analysis using stratified Cox models adjusting for multiple testing. In 2424 patients (313 PCa deaths), we evaluated the association of prediagnostic circulating IGF1 and IGFBP3 levels and PCa mortality. All statistical tests were two-sided.

**Results** The IGF signaling pathway was associated with PCa mortality ( $P = .03$ ), and *IGF2-AS* and *SSTR2* were the main contributors (both  $P = .04$ ). In SNP-specific analysis, 36 SNPs were associated with PCa mortality with  $P_{\text{trend}}$  less than .05, but only three SNPs in the *IGF2-AS* remained statistically significant after gene-based corrections. Two were in linkage disequilibrium ( $r^2 = 1$  for rs1004446 and rs3741211), whereas the third, rs4366464, was independent ( $r^2 = 0.03$ ). The hazard ratios (HRs) per each additional risk allele were 1.19 (95% confidence interval [CI] = 1.06 to 1.34;  $P_{\text{trend}} = .003$ ) for rs3741211 and 1.44 (95% CI = 1.20 to 1.73;  $P_{\text{trend}} < .001$ ) for rs4366464. rs4366464 remained statistically significant after correction for all SNPs ( $P_{\text{trend,corr}} = .04$ ). Prediagnostic IGF1 (HR<sub>highest vs lowest quartile</sub> = 0.71; 95% CI = 0.48 to 1.04) and IGFBP3 (HR = 0.93; 95% CI = 0.65 to 1.34) levels were not associated with PCa mortality.

**Conclusions** The IGF signaling pathway, primarily *IGF2-AS* and *SSTR2* genes, may be important in PCa survival.

JNCI J Natl Cancer Inst (2014) 106(5): dju085

Abundant experimental evidence indicates that the insulin-like growth factor (IGF) signaling pathway is important for cell survival and tumorigenesis (1,2). Epidemiological research, focused primarily on IGF1 and IGF binding protein 3 (IGFBP3) and risk of incident prostate cancer, suggests that higher circulating IGF1 were associated with increased risk of prostate cancer (3), with mixed findings for IGFBP3 levels (4). However, little is known about the role of prediagnostic circulating levels of IGF1 and/or IGFBP3 in prostate cancer survival.

Data on genetic variations in IGF-related genes and prostate cancer survival are sparse, limited by relatively small number of fatal outcomes and assessment of only a handful of single nucleotide polymorphisms (SNPs) related to risk of prostate cancer, as identified by tagging SNPs or from genome-wide association studies (5,6). To the best of our knowledge, a systematic evaluation of

genetic variants of IGF pathway-related genes and progression to fatal prostate cancer is lacking.

The National Cancer Institute Breast and Prostate Cancer Cohort Consortium (BPC3), pooled data from multiple large cohort studies, was designed to examine associations of variations in genes that mediate the steroid hormone and the IGF signaling pathway with breast and prostate cancer risk (7). With an average 8.9 years of follow-up among 5887 prostate cancer patients of European ancestry in BPC3, we aimed to 1) use a novel kernel machine pathway analysis and SNP-specific analysis to evaluate whether common variations among 26 genes involved in the synthesis, metabolism, and regulation of IGFs were associated with prostate cancer mortality; and 2) investigate the associations of prediagnostic circulating IGF1 and IGFBP3 levels with prostate cancer mortality in a subset of 2424 patients.

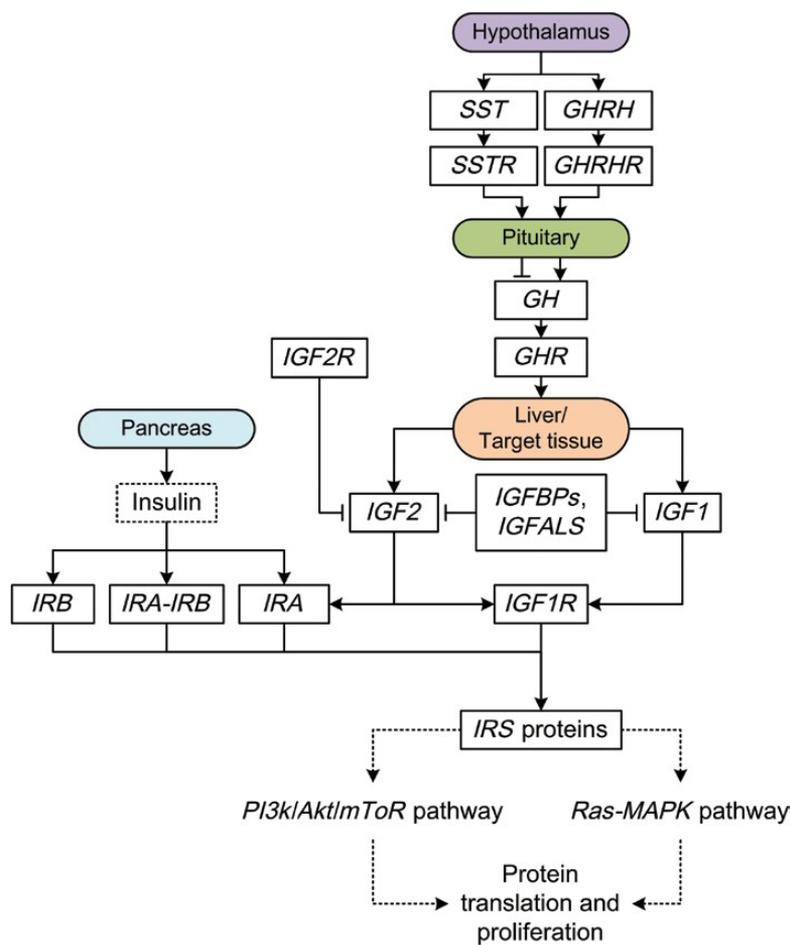
## Methods

### Study Population

The BPC3 consists of seven nested case-control studies of prostate cancer from prospective cohort studies in the United States and Europe: Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study (ATBC), American Cancer Society Cancer Prevention Study II (CPS-II), European Prospective Investigation into Cancer and Nutrition (EPIC), Health Professionals Follow-up Study (HPFS), Multiethnic Cohort Study (MEC), Physicians' Health Study (PHS), and Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial (7). Prostate cancer case patients were ascertained through population-based registries, self-report, or death certificates and verified by medical records. Height, body weight, and family history of prostate cancer were obtained by self-report. Data on disease stage (Jewett-Whitmore classification) and grade (Gleason score) were collected from each cohort. Written informed consent was obtained from all subjects, and each study was approved by the institutional review boards at their respective institutions. Details of vital status follow-up and determination of cause of death are described in the [Supplementary Methods](#) (available online).

### SNP Selection and Genotyping

A total of 590 SNPs in 26 genes involved in the synthesis, metabolism, and regulation of insulin-like growth factors were genotyped (Figure 1). After restricting to self-reported European ancestry, a total of 5887 prostate cancer patients were included in this analysis. Two approaches were taken to evaluate linkage disequilibrium (LD) patterns and select the SNPs for this analysis as described elsewhere (7,8). Genotyping was performed in six laboratories: National Cancer Institute Core Genotyping Facility (Gaithersburg, MD), University of Southern California (Los Angeles, CA), University of Hawaii (Honolulu, HI), Harvard School of Public Health (Boston, MA), Imperial College (London, UK), and Cambridge University (Cambridge, UK). A total of 40 SNPs from *GNRH1*, *GNRHR*, *IGF1*, *IGFBP1*, and *IGFBP3* were genotyped using TaqMan (Applied Biosystems, Foster City, CA). The remaining SNPs were genotyped by Illumina Golden Gate platform (San Diego, CA). Interlaboratory concordance was evaluated by genotyping 94 samples from the SNP 500 cancer panel (9) for the TaqMan SNPs and 30 HapMap CEU (Utah residents with ancestry from northern and western Europe) trios for the Illumina panel, with concordance rates greater than 99% between laboratories.



**Figure 1.** IGF signaling pathway. Genes included in this analysis were *SST*, *SSTR1-5*, *GHRH*, *GHRHR*, *GH*, *IGF1*, *IGF1R*, *IGFBP1-6*, *IGF2-AS*, *IGF2R*, *IGFBPs*, *IGFBPs*, *IGFALS*, *INSR*, *IRS1*, *IRS2* (shown in Figure 1) and *POU1F1*, *GNRH1*, and *GNRHR* (not shown); the insulin receptor is encoded by a single gene, *INSR*, from which alternate splicing during transcription results in either *IRA* or *IRB* isoforms; the insulin gene (*INS*) was not genotyped, and genes in PI3k/Akt/mTOR and Ras-MAPK pathway were not included in this analysis.

Genotype data from the Taqman and Illumina platforms were filtered separately. Any sample in which more than 25% of the SNPs attempted on a given platform failed was removed from the dataset. Within each study, any SNP that failed in 25% or more of the samples, exhibited a statistically significant ( $P < 10^{-5}$ ) deviation from Hardy–Weinberg proportions among European-ancestry controls, or had a minor allele frequency less than 1% was removed from the dataset. SNPs that were missing in more than 25% of the prostate cancer patients or showed large differences in allele frequency among subjects with European ancestry across studies (fixation index  $F_{st} > 0.02$ ) were also excluded from analysis. For each gene region, SNPs that were polymorphic in any of the HapMap reference panels were imputed using MACH (10). Genotypes were imputed by cohort using the CEPH (Utah residents with ancestry from northern and western Europe) European (CEU) reference panel for subjects of European ancestry (release No. 21). Imputed data was filtered by study, and poorly imputed SNPs ( $r^2 < 0.3$ ) were removed from analysis.

### Circulating IGF1 and IGFBP3 levels

Prediagnostic measurements of IGF1 and IGFBP3 were available for five of the seven cohorts (ATBC, EPIC, HPFS, PHS, and PLCO;  $n = 2445$ ) (11–15). Details of sample collection and storage were described previously. Samples from ATBC, HPFS, and PHS were measured in the Pollak laboratory (McGill University, Montreal, QC, Canada), and the remaining studies were measured in the laboratory of the Hormones and Cancer Team at International Agency for Research on Cancer (IARC) with enzyme-linked immunosorbent assays (Diagnostic System Laboratories, Webster, TX). We excluded cohort and assay batch-specific statistical outliers ( $n = 21$ ) based on the generalized extreme studentized deviate many-outlier detection approach, setting alpha to 0.05 for both IGF1 and IGFBP3 blood levels (16).

### Statistical Analysis

**IGF Gene Pathway.** The kernel machine Cox regression framework (17,18), a novel and comprehensive approach for pathway analysis of censored survival outcomes, was used to assess associations with deaths from prostate cancer and other causes for SNP sets defined by all 26 genes in the IGF pathway and each gene individually after adjusting for continuous age and study cohort. Because genotyped SNPs may be imperfect surrogates for the true causal SNP, their individual relative risks are likely to be modest, and a multimarker global test will more effectively capture the true effect. The kernel machine accounts for LD in an SNP set, leading to a powerful test with reduced degrees of freedom. More attractively, it can also capture potential nonlinear SNP effects, SNP–SNP interactions (epistasis), and the joint effects of multiple causal variants without requiring a priori knowledge of directionality. The kernel machine tests whether an SNP set is associated with event time of interest after adjusting for covariables, and the test statistic under the null follows a mixture of  $\chi^2$  distributions, which can be approximated by resampling methods. Logistic kernel machines have been applied in a variety of traits and diseases (19,20).

SNP-specific analyses were conducted by stratified Cox proportional hazards models under a log-additive hazards assumption and stratified by study cohort, allowing different baseline hazards for each study. Follow-up was defined from the date of prostate cancer diagnosis to the date of any death or last follow-up. The

assumption of proportionality was verified by testing each SNP and time since diagnosis, and no violation was identified. All analyses were adjusted for age at diagnosis and further adjusted for stage and Gleason score at diagnosis. To correct for multiple testing with possible presence of LD, the number of effective SNPs,  $M_{\text{eff}}$ , was calculated for each gene using a spectral decomposition approach (21). For gene-based  $P$  value correction, nominal  $P$  values for each SNP were multiplied by the  $M_{\text{eff}}$  for the gene. For the pathway-based correction, the  $M_{\text{eff}}$  values for all 26 genes were summed to correct the  $P$  values.

Cumulative incidence of prostate cancer death by years since diagnosis were plotted for statistically significant SNPs after gene-level-based correction using competing-risks regression by the method of Fine and Gray (22).

Stratified analysis of statistically significant SNPs and prostate cancer mortality association by age at diagnosis ( $<65$  or  $\geq 65$  years) and BMI ( $<25$ ,  $25\text{--}30$ , or  $\geq 30$  kg/m<sup>2</sup>), Gleason score (2–6, 7, or 8–10) and stage (A/B or C/D) were conducted under a dominant model as a result of limited sample size. To assess effect modification, we added a product term of statistically significant SNPs with the variables above and computed  $P$  values from log likelihood ratio test.

**Circulating IGF1 and IGFBP3 Levels.** We created batch-specific ( $n = 10$ ) quartiles for IGF1 and IGFBP3 and assessed their associations with prostate cancer mortality simultaneously by stratified Cox proportional hazards models adjusting for age at diagnosis. Models were also additionally adjusted for BMI assessed at the baseline of each study to assess possible confounding or stage and Gleason score at diagnosis to evaluate possible mediation. Tests for trend were done by treating the median concentration for each quartile as a continuous variable. Stratified analysis by stage and Gleason score at diagnosis were also performed. To account for the possibility of reverse causation in which an undiagnosed tumor could affect biomarker levels, sensitivity analyses were conducted by excluding cases diagnosed within 2 years of blood draw.

Analyses were conducted using SAS 9.2 (SAS Institute, Cary, NC), R (The R Foundation for Statistical Computing; <http://www.r-project.org/foundation/>), and Stata 12 (StataCorp, College Station, TX). All statistical tests were two-sided. A  $P$  value of less than .05 was considered statistically significant.

## results

During an average follow-up of 8.9 years among the 5887 case patients, 1,999 patients died, 704 of whom had prostate cancer as the underlying cause of death. Among the 2424 men in the subgroup of biomarker analysis, 313 of the 810 deaths were due to prostate cancer. Compared with those who were either alive at last follow-up or had died from other causes, patients who died from prostate cancer had higher Gleason score and clinical stage at diagnosis but similar BMI (Table 1; Supplementary Table 1, available online).

### IGF Gene Pathway and Prostate Cancer Mortality

**Pathway Analysis.** A total of 530 SNPs were included in the genetic analysis. Kernel machine pathway analysis suggests that this set of SNPs covering all 26 genes in the IGF signaling pathway

**Table 1.** Characteristics of prostate cancer patients in the National Cancer Institute Breast and Prostate Cancer Cohort Consortium\*

Characteristic	PCa death (n = 704)	Censored (n = 5183)	Total (n = 5887)
Age at diagnosis, y, mean (SD)	69.1 (7.1)	68.3 (6.4)	68.4 (6.5)
Diagnosis to prostate cancer death/ censoring, y, mean (SD)	5.3 (3.8)	9.4 (3.9)	8.9 (4.1)
Body mass index, kg/m <sup>2</sup>			
18–24.9	265 (38)	2030 (39)	2295 (39)
25–29.9	342 (49)	2393 (46)	2735 (46)
≥30	78 (11)	572 (11)	650 (11)
Missing	19 (2)	188 (4)	207 (4)
Family history			
Yes	39 (6)	576 (11)	615 (10)
No	358 (51)	3038 (59)	3396 (58)
Missing	307 (44)	1569 (30)	1876 (32)
Gleason score			
2–6	115 (16)	2567 (50)	2682 (46)
7	225 (32)	1465 (28)	1690 (29)
8–10	217 (31)	485 (9)	702 (12)
Missing	147 (21)	666 (13)	813 (14)
Stage			
A or B	259 (37)	3801 (73)	4060 (69)
C or D	343 (49)	702 (14)	1045 (18)
Missing	102 (14)	680 (13)	782 (13)
Biomarker subcohort			
No. of patients	313	2111	2424
Age at blood draw, y, mean (SD)	64.0 (7.8)	63.0 (6.9)	63.1 (7.1)
Circulating IGF1, ng/mL, median (IQR)	161 (124–212)	182 (142–228)	179 (139–227)
Circulating IGFBP3, ng/mL, median (IQR)	3110 (2544–3753)	3613 (2597–4333)	3544 (2899–4290)

\* Data are No. (%) unless otherwise specified.

was associated with prostate cancer mortality ( $P = .03$ ) (Table 2). When testing the SNP set of each gene, *IGF2-AS* (9 SNPs;  $P = .04$ ) and *SSTR2* (14 SNPs;  $P = .04$ ) showed statistically significant associations with prostate cancer mortality. The overall pathway  $P$  values were .05 without either *IGF2-AS* or *SSTR2* and .08 without both *IGF2-AS* and *SSTR2*, suggesting both *IGF2-AS* and *SSTR2* may contribute to the progression to fatal prostate cancer. Neither the overall pathway nor *IGF2-AS* or *SSTR2* were associated with risk of dying from causes other than prostate cancer.

**SNP-Specific Analysis.** A total of 36 SNPs were associated with prostate cancer mortality with  $P_{\text{trend}} < .05$  (Supplementary Table 2, available online). After correcting for multiple testing at gene level, three SNPs, all in *IGF2* antisense gene (*IGF2-AS*, 11p15.5), were statistically significantly associated with prostate cancer-specific mortality. Two of these SNPs, rs1004446 (intron) and rs3741211(3'-UTR), were in LD with each other ( $r^2 = 1$  in 1000 Genome CEU population) but independent with the third SNP rs4366464 (intron) ( $r^2 = 0.03$ ). For rs3741211, each additional A allele was associated with a 19% (hazard ratio [HR] = 1.19; 95% confidence interval [CI] = 1.06 to 1.34;  $P_{\text{trend}} = .003$ ) increased risk of prostate cancer-specific mortality. For rs4366464, each additional minor allele G was associated with a 44% (HR = 1.44; 95% CI = 1.20 to 1.73) increased risk of prostate cancer mortality ( $P_{\text{trend}} = <.001$ ) (Table 3; Supplementary Figure 1, available online). The association for rs4366464 remained statistically significant after further correcting for multiple testing of all SNPs ( $P_{\text{trend,corr}} = .04$ ;  $M_{\text{eff}} = 424$ ). When mutually adjusted for each other, the hazard ratios remained similar for rs3741211 (HR = 1.15; 95% CI = 1.03 to 1.30) and rs4366464 (HR = 1.37; 95% CI = 1.13 to

1.67), suggesting independent additive effects of the two SNPs on prostate cancer progression. Cohort-specific associations (Figure 2) also indicated the robustness of these associations, and minimal heterogeneities were observed (rs3741211:  $I^2 < 0.05\%$ ,  $P_{\text{heterogeneity}} = .44$ ; rs4366464:  $I^2 < 0.05\%$ ,  $P_{\text{heterogeneity}} = .55$ ).

SNP rs4366464 or rs3741211 was not statistically significantly associated with either Gleason score or stage (data not shown). After additionally adjusting for these clinical parameters, the association between rs3741211 and prostate cancer death remained unchanged, whereas the hazard ratio for rs4366464 was slightly attenuated. Neither rs3741211 nor rs4366464 was associated with risk of dying from other causes (Table 3). These data suggest that the association between the two SNPs in *IGF2-AS* and prostate cancer mortality were independent of tumor characteristics and specific to death from prostate cancer.

Joint effect analysis suggests that for rs3741211, the association with prostate cancer mortality tended to be stronger among men with cancer diagnosed at younger age or patients with BMI less than 25 kg/m<sup>2</sup> (Supplementary Figure 2, available online). For rs4366464, the association was stronger among men diagnosed at younger age. For both SNPs, the associations were somewhat stronger among patients with higher stage (C or D) or higher Gleason score ( $\geq 7$ ). However, only interaction between rs3741211 and stage was statistically significant ( $P = .02$ ).

### Circulating IGF1 and IGFBP3 and Prostate Cancer Mortality

IGF1 levels were statistically significantly correlated with IGFBP3 ( $r = 0.52$ ;  $P < .001$ ). Prediagnostic circulating levels of IGF1 (HR<sub>highest vs lowest quartile</sub> = 0.71; 95% CI = 0.48 to 1.04) and IGFBP3 (HR = 0.93; 95% CI = 0.65 to 1.34) were not associated with prostate cancer

**Table 2.** IGF pathway analyses for prostate cancer–specific mortality and mortality of other causes by kernel machine\*

Gene abbreviation	Gene name	Chromosomal region	No. of SNPs included	<i>P</i> for PCa death†	<i>P</i> for other death†
Pathway					
Total pathway	—	—	530	.03	.14
Pathway w/o <i>IGF2-AS</i>	—	—	521	.05	.14
Pathway w/o <i>SSTR2</i>	—	—	516	.05	.13
Pathway w/o <i>IGF2-AS</i> and <i>SSTR2</i>	—	—	507	.08	.13
Gene					
<i>GHR</i>	Growth hormone receptor	5p13-p12	34	.16	.61
<i>GHRH</i>	Growth hormone releasing hormone	20q11.2	9	.14	.75
<i>GHRHR</i>	Growth hormone releasing hormone receptor	7p14	26	.38	.85
<i>GNRHI</i>	Gonadotropin-releasing hormone 1	8p21-p11.2	3	.86	.55
<i>GNRHR</i>	Gonadotropin-releasing hormone receptor	4q21.2	6	.15	.07
<i>IGF1</i>	Insulin-like growth factor 1	12q23.2	14	.35	.12
<i>IGF1R</i>	Insulin-like growth factor 1 receptor	15q26.3	112	.36	.20
<i>IGF2-AS</i>	IGF2 antisense RNA	11p15.5	9	.04	.40
<i>IGF2R</i>	Insulin-like growth factor 2 receptor	6q26	68	.09	.16
<i>IGFALS</i>	Insulin-like growth factor binding protein, acid labile subunit	16p13.3	7	.34	.54
<i>IGFBP1</i>	Insulin-like growth factor binding protein 1	7p13-p12	7	.29	.22
<i>IGFBP2,5</i>	Insulin-like growth factor binding protein 2 and 5	2q33-q36	36	.06	.34
<i>IGFBP3</i>	Insulin-like growth factor binding protein 3	7p13-p12	8	.77	.22
<i>IGFBP4</i>	Insulin-like growth factor binding protein 4	17q12-q21.1	7	.69	.67
<i>IGFBP6</i>	Insulin-like growth factor binding protein 6	12q13	7	.76	.54
<i>INSR</i>	Insulin receptor	19p13.3-p13.2	53	.07	.19
<i>IRS1</i>	Insulin receptor substrate 1	2q36	8	.53	.85
<i>IRS2</i>	Insulin receptor substrate 2	13q34	13	.38	.82
<i>POU1F1</i>	POU class 1 homeobox 1	3p11	6	.55	.42
<i>SST</i>	Somatostatin	3q28	16	.66	.03
<i>SSTR1</i>	Somatostatin receptor 1	14q13	19	.31	.06
<i>SSTR2</i>	Somatostatin receptor 2	17q24	14	.04	.50
<i>SSTR3</i>	Somatostatin receptor 3	22q13.1	18	.96	.96
<i>SSTR4</i>	Somatostatin receptor 4	20p11.2	26	.24	.41
<i>SSTR5</i>	Somatostatin receptor 5	16p13.3	4	.78	.91

\* PCa = prostate cancer; SNP = single nucleotide polymorphism.

† *P* values were calculated using kernel machine Cox regression framework and were two-sided.

mortality in the model mutually adjusted for each other and age at diagnosis (Table 4). The hazard ratios were similar after additionally adjusting for stage and Gleason score at diagnosis in the model, or BMI at baseline, or excluding IGF1 and IGFBP3 measurements within 2 years of prostate cancer diagnosis (data not shown). In subgroup analysis, higher IGF1 levels were statistically significantly associated with lower prostate cancer mortality ( $P_{\text{trend}} = .02$ ) among men diagnosed with more advanced tumors (stage C or D).

## Discussion

To the best of our knowledge, this analysis of IGF pathway genes in relation to prostate cancer mortality among prostate cancer patients is the largest study to date. Using the kernel machine pathway analysis, a powerful test allowing assessment of the joint associations of variants in a predefined pathway, we demonstrated that the IGF pathway was statistically significantly associated with prostate cancer mortality and two genes, *IGF2-AS* and *SSTR2*, may play important roles in prostate cancer progression. Using SNP-specific association analysis, we further identified two SNPs, rs3741211 and rs4366464 in *IGF2-AS*, that were statistically significantly associated with prostate cancer mortality.

Additionally, among a subset of 2424 patients, we found no overall associations between prediagnostic circulating levels of IGF1 and IGFBP3 and prostate cancer mortality. The null associations between *IGF1* and *IGFBP3* genes and prostate cancer mortality suggest that their roles in the progression of prostate cancer were limited. In previous analyses of BPC3 patients, genetic variations in *IGF1* and *SSTR5* were associated with circulating levels of IGF1, and *IGFBP3* and *IGFALS* genes were associated with IGFBP3 levels (8,23). However, none of the SNPs in these genes were associated with prostate cancer mortality in our analysis, which is in line with the null findings between circulating levels of IGF1 and IGFBP3 and prostate cancer mortality. Although these findings should be interpreted with caution given the heterogeneities in blood collection, sample storage, and assay variation across the cohorts, the findings are not surprising because recent prospective studies did not support stronger associations of IGF1 levels with risk of advanced prostate cancer, favoring the hypothesis that common germline variations or circulating levels of IGF1 may contribute to early growth of prostate carcinogenesis (4), but not during progression.

The role of *IGF2-AS* and *IGF2* in prostate cancer initiation and progression is largely underexplored. A previous genome-wide

**Table 3.** Single nucleotide polymorphisms in *IGF2-AS* associated with prostate cancer–specific mortality after gene-based *P* value correction\*

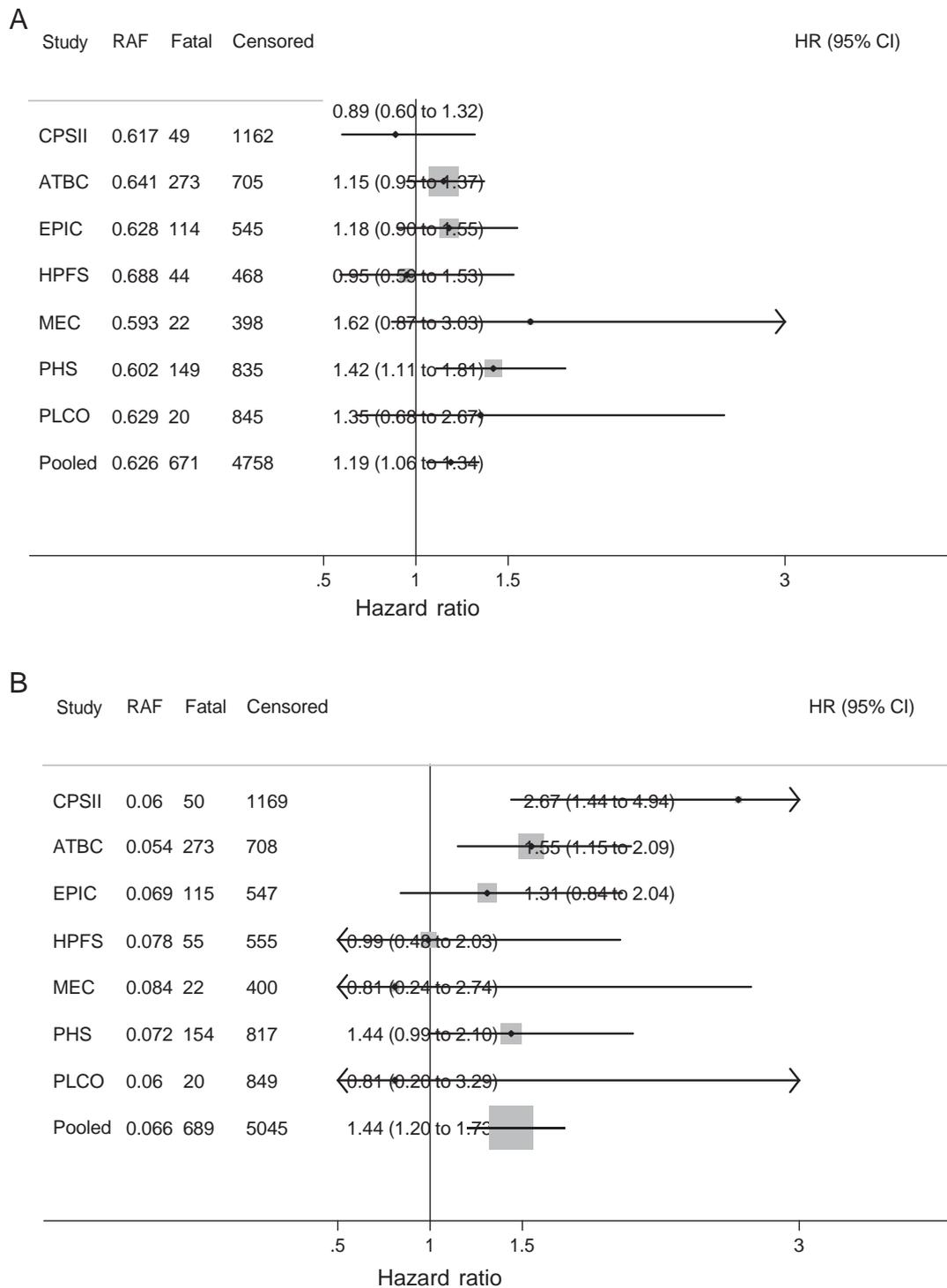
SNP	Risk allele	RAF	Chromosomal region	Position	Genotype	Person-years	PCa death			Other death		
							No.	HR (95% CI)†	HR (95% CI)‡	No.	HR (95% CI)†	HR (95% CI)‡
rs3741211	A	0.626	11p15.5	2169110	GG	6945	64	1.00 (referent)	1.00 (referent)	182	1.00 (referent)	1.00 (referent)
					GA	23 425	312	1.40 (1.07 to 1.83)	1.43 (1.09 to 1.87)	558	0.89 (0.75 to 1.05)	0.89 (0.75 to 1.05)
					AA	19 562	295	1.55 (1.18 to 2.03)	1.50 (1.15 to 1.97)	500	0.93 (0.79 to 1.11)	0.93 (0.78 to 1.10)
					AA/GA	42 987	607	1.47 (1.13 to 1.90)	1.46 (1.13 to 1.90)	1058	0.91 (0.78 to 1.07)	0.90 (0.77 to 1.06)
					per allele	—	—	1.19 (1.06 to 1.34)	1.16 (1.04 to 1.30)	—	0.99 (0.91 to 1.07)	0.98 (0.90 to 1.07)
					<i>P</i> <sub>trend</sub> §	—	—	.003	.01	—	.75	.66
					<i>P</i> <sub>trend,corr</sub> §	—	—	.02	.08	—	1.00	1.00
rs4366464	G	0.066	11p15.5	2164799	CC	44 133	566	1.00 (referent)	1.00 (referent)	1099	1.00 (referent)	1.00 (referent)
					GC	6596	117	1.39 (1.14 to 1.70)	1.32 (1.08 to 1.62)	157	0.96 (0.82 to 1.14)	0.96 (0.81 to 1.14)
					GG	169	6	2.87 (1.28 to 6.44)	2.34 (1.04 to 5.25)	7	1.93 (0.92 to 4.08)	1.88 (0.89 to 3.97)
					GG/GC	6764	123	1.43 (1.18 to 1.74)	1.35 (1.11 to 1.65)	164	0.99 (0.84 to 1.16)	0.98 (0.83 to 1.16)
					per allele	—	—	1.44 (1.20 to 1.73)	1.36 (1.13 to 1.63)	—	1.01 (0.86 to 1.18)	1.00 (0.86 to 1.18)
					<i>P</i> <sub>trend</sub> §	—	—	.0001	.001	—	.92	.96
					<i>P</i> <sub>trend,corr</sub> §	—	—	.0008	.01	—	1.00	1.00

\* CI = confidence interval; HR = hazard ratio; PCa = prostate cancer; SNP = single nucleotide polymorphism; RAF = risk allele frequency in patients who did not die from prostate cancer.

† The Cox model was stratified by study cohort and adjusted for age at diagnosis.

‡ The model was additionally adjusted for Gleason score and stage at diagnosis. Because adding body mass index to the multivariable model did not alter the hazard ratios, we decided not to present results adjusted for body mass index.

§ *P*<sub>trend</sub> were calculated using stratified Cox proportional hazards models under a log-additive hazards assumption and were two-sided. *P*<sub>trend,corr</sub> were *P*<sub>trend</sub> after gene-based correction for multiple testing (*M*<sub>eff</sub> = 8).



**Figure 2.** Association of *IGF2-AS* single nucleotide polymorphism rs3741211 and rs4366464 with prostate cancer-specific mortality by study cohort. Hazard ratios (HRs; **diamonds**) and 95% confidence intervals (CIs; **error bars**) calculated for the association for the individual studies and the pooled analysis for rs3741211 (**A**) and rs4366464 (**B**) are shown. **Size of gray square** represents percentage weight of each study. RAF = risk allele

frequency. ATBC = Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study; CI = confidence interval; CPS-II = American Cancer Society Cancer Prevention Study II; EPIC = European Prospective Investigation into Cancer and Nutrition; HPFS = Health Professionals Follow-up Study; HR = hazard ratio; MEC = Multiethnic Cohort Study; PHS = Physicians' Health Study; PLCO = Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial.

association study identified SNP rs7127900 in *IGF2-AS* as associated with risk of incident prostate cancer (24) but not with prostate cancer mortality (5). This SNP was not in LD with the two SNPs we identified ( $r^2 = 0.01$  for rs3741211 and  $r^2 = 0.003$  for rs4366464 in 1000 Genome CEU population).

IGF2 is a peptide growth factor that is homologous to both IGF1 and insulin; interaction of IGF2 with insulin receptor sub-type A (IRA) may play a role both in fetal growth and cancer biology (25). *IGF2-AS* expresses a paternally imprinted antisense transcript of the *IGF2* gene. It is transcribed in the opposite

**Table 4.** Circulating levels of IGF1 and IGFBP3 and prostate cancer–specific mortality in the National Cancer Institute Breast and Prostate Cancer Cohort Consortium

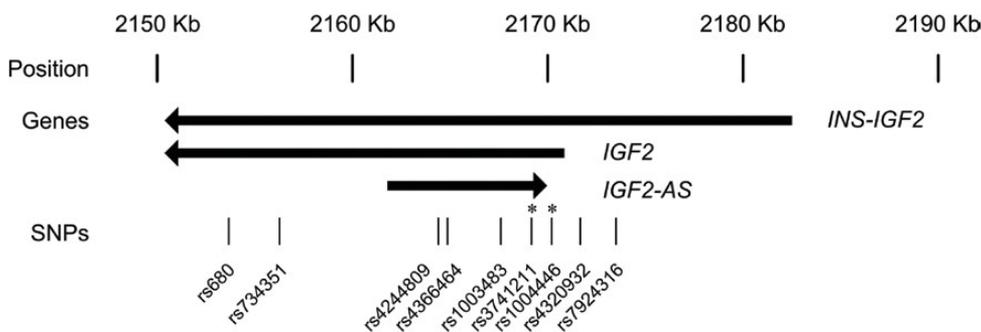
Outcome/ biomarker	Quartile*								<i>P</i> <sub>trend</sub> §
	Q1		Q2		Q3		Q4		
	Fatal/ censored	HR(95% CI)	Fatal/ censored	HR(95% CI)	Fatal/ censored	HR(95% CI)	Fatal/ censored	HR(95% CI)	
All cases									
Model 1†									
IGF1	101/501	1.00 (referent)	80/529	0.86 (0.63 to 1.17)	68/541	0.74 (0.53 to 1.05)	64/540	0.71 (0.48 to 1.04)	.08
IGFBP3	102/499	1.00 (referent)	77/534	0.83 (0.61 to 1.13)	59/549	0.67 (0.47 to 0.95)	75/529	0.93 (0.65 to 1.34)	.35
Model 2‡									
IGF1	101/501	1.00 (referent)	80/529	0.84 (0.62 to 1.14)	68/541	0.77 (0.55 to 1.09)	64/540	0.77 (0.52 to 1.14)	.18
IGFBP3	102/499	1.00 (referent)	77/534	0.77 (0.57 to 1.06)	59/549	0.59 (0.41 to 0.84)	75/529	0.93 (0.65 to 1.35)	.28
Stage A or B†									
IGF1	35/357	1.00 (referent)	23/375	0.81 (0.46 to 1.41)	27/363	1.05 (0.59 to 1.89)	19/374	0.75 (0.37 to 1.53)	.53
IGFBP3	39/347	1.00 (referent)	25/376	0.65 (0.37 to 1.12)	13/381	0.37 (0.19 to 0.74)	27/365	0.77 (0.41 to 1.46)	.23
Stage C or D†									
IGF1	48/56	1.00 (referent)	38/80	0.73 (0.46 to 1.16)	32/91	0.58 (0.35 to 0.94)	30/89	0.52 (0.30 to 0.90)	.02
IGFBP3	40/64	1.00 (referent)	40/82	0.99 (0.63 to 1.55)	34/83	0.91 (0.55 to 1.52)	34/87	1.26 (0.73 to 2.19)	.38
Gleason <7†									
IGF1	19/246	1.00 (referent)	13/267	0.72 (0.34 to 1.53)	14/272	0.78 (0.36 to 1.70)	12/297	0.68 (0.28 to 1.68)	.49
IGFBP3	19/247	1.00 (referent)	15/274	0.86 (0.42 to 1.79)	11/283	0.68 (0.30 to 1.55)	13/278	0.85 (0.36 to 2.02)	.64
Gleason ≥7†									
IGF1	56/174	1.00 (referent)	48/180	0.93 (0.62 to 1.41)	30/188	0.64 (0.39 to 1.06)	35/172	0.81 (0.47 to 1.40)	.33
IGFBP3	57/169	1.00 (referent)	34/182	0.72 (0.46 to 1.12)	40/179	0.85 (0.54 to 1.34)	38/184	0.83 (0.49 to 1.41)	.63

\* Batch-specific (n = 10) quartiles were used. All models were stratified by study cohort and simultaneously adjusted for IGF1 and IGFBP3. CI = confidence interval; HR = hazard ratio.

† Adjusted for age at diagnosis

‡ Adjusted for age, Gleason score, and stage at diagnosis

§ *P*<sub>trend</sub> values were calculated by treating the median concentration for each quartile as a continuous variable and were two-sided.



**Figure 3.** Gene map of *IGF2-AS/IGF2/INS* region and single nucleotide polymorphisms (SNPs) genotyped in *IGF2-AS* (n = 9). Only SNPs rs1004446 and rs3741211 have an *r*<sup>2</sup> greater than 0.8, indicated by an asterisk (\*).

direction to the *IGF2* transcripts, with some genomic regions shared with *IGF2* (Figure 3) (26). *IGF2-AS* and *IGF2* were overexpressed in Wilms' tumor through loss of imprinting (26,27). Loss of imprinting of *IGF2* is generally manifested by the activation of the normally silenced maternal allele with the subsequent expression of both gene copies. Evidence from Wilms' tumor, colorectal cancer, and ovarian cancer suggests that the biallelic *IGF2* expression also correlates with aberrant *IGF2/H19* methylation (28,29). IGF2 levels were increased in prostate tumor-associated tissues, and a widespread *IGF2* loss of imprinting throughout the peripheral prostate in men with prostate cancer was observed but not in samples of benign prostatic hyperplasia or other adult tissues, suggesting that epigenetic modification may play an important role in prostate cancer carcinogenesis

(30). Overexpression of *IGF2* and/or *IRA* has been proposed as a potential mechanism of resistance to IGF1R-directed therapies (31).

SSTR2 has been documented in experimental and clinical prostate cancer research but not in population studies. Somatostatin exerts inhibitory effects on cancer cells, including prostate, through five specific G-protein-coupled membrane receptors, SSTR1–5, with SSTR2 being predominant in human cancers (32,33). Its analogs, octreotide and lanreotide, which have high affinity for SSTR2, have been used to treat hormone-refractory prostate cancers (34,35) but are still under development.

The major strength of this study is the use of a large cohort consortium to study genetic predispositions, which are less likely to be affected by screening and treatment. Another strength is our

comprehensive evaluations of genetic variants in the IGF pathway using pathway, SNP-specific, and study cohort-specific analysis. However, additional genotyping to narrow down the region harboring the causal allele, followed by functional work on the identified variants and validations in other independent studies and/or races/ethnicities are necessary. Lack of patient treatment information was another limitation. However, associations of IGF genetic polymorphisms or biomarkers with prostate cancer mortality were unlikely to be affected by treatment because the two SNPs we identified, rs3741211 and rs4366464, were not associated with tumor characteristics (stage and Gleason score), the major determinants of treatment.

In summary, in this large consortium analysis of prostate cancer, both pathway and SNP-specific analyses showed that germline variations in *IGF2-AS* gene were associated with prostate cancer mortality, independent of stage and Gleason score and specific to prostate cancer. In contrast, neither genetic polymorphisms nor prediagnostic circulating levels of IGF1 and IGFBP3 were associated with prostate cancer mortality. Pathway analysis suggests that *SSTR2* may also play a role in prostate cancer progression, but SNP-specific analysis failed to show any statistically significant SNP in this gene after gene-level correction. Further research on the role of *IGF2/IGF2-AS* and *SSTR2* is needed.

## references

- Pollak M. Insulin and insulin-like growth factor signalling in neoplasia. *Nat Rev Cancer*. 2008;8(12):915–928.
- Furstenberger G, Senn HJ. Insulin-like growth factors and cancer. *Lancet Oncol*. 2002;3(5):298–302.
- Rowlands MA, Gunnell D, Harris R, et al. Circulating insulin-like growth factor peptides and prostate cancer risk: a systematic review and meta-analysis. *Int J Cancer*. 2009;124(10):2416–2429.
- Nimptsch K, Giovannucci E. Epidemiology of IGF-1 and Cancer, *Insulin-like Growth Factors and Cancer: From Basic Biology to Therapeutics*, Cancer Drug Discovery and Development. In: LeRoith D, ed. New York, NY: Springer Science+Business Media, LLC; 2012:1–24.
- Lin DW, FitzGerald LM, Fu R, et al. Genetic variants in the *LEPR*, *CRY1*, *RNASEL*, *IL4*, and *ARVCF* genes are prognostic markers of prostate cancer-specific mortality. *Cancer Epidemiol Biomarkers Prev*. 2011;20(9):1928–1936.
- Johansson M, McKay JD, Rinaldi S, et al. Genetic and plasma variation of insulin-like growth factor binding proteins in relation to prostate cancer incidence and survival. *Prostate*. 2009;69(12):1281–1291.
- Hunter DJ, Riboli E, Haiman CA, et al. A candidate gene approach to searching for low-penetrance breast and prostate cancer genes. *Nat Rev Cancer*. 2005;5(12):977–985.
- Gu F, Schumacher FR, Canzian F, et al. Eighteen insulin-like growth factor pathway genes, circulating levels of IGF-I and its binding protein, and risk of prostate and breast cancer. *Cancer Epidemiol Biomarkers Prev*. 2010;19(11):2877–2887.
- Packer BR, Yeager M, Burdett L, et al. SNP500Cancer: a public resource for sequence validation, assay development, and frequency analysis for genetic variation in candidate genes. *Nucleic Acids Res*. 2006;34(database issue):D617–D621.
- Li Y, Willer C, Sanna S, et al. Genotype imputation. *Annu Rev Genomics Hum Genet*. 2009;10:387–406.
- Chan JM, Stampfer MJ, Ma J, et al. Insulin-like growth factor-I (IGF-I) and IGF binding protein-3 as predictors of advanced-stage prostate cancer. *J Natl Cancer Inst*. 2002;94(14):1099–1106.
- Platz EA, Pollak MN, Leitzmann MF, et al. Plasma insulin-like growth factor-I and binding protein-3 and subsequent risk of prostate cancer in the PSA era. *Cancer Causes Control*. 2005;16(3):255–262.
- Weiss JM, Huang WY, Rinaldi S, et al. IGF-1 and IGFBP-3: Risk of prostate cancer among men in the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial. *Int J Cancer*. 2007;121(10):2267–2273.
- Woodson K, Tangrea JA, Pollak M, et al. Serum insulin-like growth factor I: tumor marker or etiologic factor? A prospective study of prostate cancer among Finnish men. *Cancer Res*. 2003;63(14):3991–3994.
- Allen NE, Key TJ, Appleby PN, et al. Serum insulin-like growth factor (IGF)-I and IGF-binding protein-3 concentrations and prostate cancer risk: results from the European Prospective Investigation into Cancer and Nutrition. *Cancer Epidemiol Biomarkers Prev*. 2007;16(6):1121–1127.
- Rosner B. Percentage points for a generalized esd many-outlier procedure. *Technometrics*. 1983;25(2):165–172.
- Lin X, Cai T, Wu MC, et al. Kernel machine SNP-set analysis for censored survival outcomes in genome-wide association studies. *Genet Epidemiol*. 2011;35(7):620–631.
- Cai T, Tonini G, Lin X. Kernel machine approach to testing the significance of multiple genetic markers for risk prediction. *Biometrics*. 2011;67(3):975–986.
- Shui IM, Mucci LA, Kraft P, et al. Vitamin D-related genetic variation, plasma vitamin D, and risk of lethal prostate cancer: a prospective nested case-control study. *J Natl Cancer Inst*. 2012;104(9):690–699.
- Liu CY, Wu MC, Chen F, et al. A large-scale genetic association study of esophageal adenocarcinoma risk. *Carcinogenesis*. 2010;31(7):1259–1263.
- Gao X, Starmer J, Martin ER. A multiple testing correction method for genetic association studies using correlated single nucleotide polymorphisms. *Genet Epidemiol*. 2008;32(4):361–369.
- Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *J Am Stat Assoc*. 1999;94(446):496–509.
- Schumacher FR, Cheng I, Freedman ML, et al. A comprehensive analysis of common IGF1, IGFBP1 and IGFBP3 genetic variation with prospective IGF-I and IGFBP-3 blood levels and prostate cancer risk among Caucasians. *Hum Mol Genet*. 2010;19(15):3089–3101.
- Eeles RA, Kote-Jarai Z, Al Olama AA, et al. Identification of seven new prostate cancer susceptibility loci through a genome-wide association study. *Nat Genet*. 2009;41(10):1116–1121.
- Frasca F, Pandini G, Scalia P, et al. Insulin receptor isoform A, a newly recognized, high-affinity insulin-like growth factor II receptor in fetal and cancer cells. *Mol Cell Biol*. 1999;19(5):3278–3288.
- Okutsu T, Kuroiwa Y, Kagitani F, et al. Expression and imprinting status of human PEG8/IGF2AS, a paternally expressed antisense transcript from the IGF2 locus, in Wilms' tumors. *J Biochem*. 2000;127(3):475–483.
- Ogawa O, Eccles MR, Szeto J, et al. Relaxation of insulin-like growth factor-II gene imprinting implicated in Wilms tumor. *Nature*. 1993;362(6422):749–751.
- Murphy SK, Huang Z, Wen Y, et al. Frequent IGF2/H19 domain epigenetic alterations and elevated IGF2 expression in epithelial ovarian cancer. *Mol Cancer Res*. 2006;4(4):283–292.
- Nakagawa H, Chadwick RB, Peltomaki P, et al. Loss of imprinting of the insulin-like growth factor II gene occurs by biallelic methylation in a core region of H19-associated CTCF-binding sites in colorectal cancer. *Proc Natl Acad Sci U S A*. 2001;98(2):591–596.
- Bhusari S, Yang B, Kueck J, et al. Insulin-like growth factor-2 (IGF2) loss of imprinting marks a field defect within human prostates containing cancer. *Prostate*. 2011;71(15):1621–1630.
- Hendrickson AW, Haluska P. Resistance pathways relevant to insulin-like growth factor-1 receptor-targeted therapy. *Curr Opin Invest Drugs*. 2009;10(10):1032–1040.
- Reubi JC. Peptide receptors as molecular targets for cancer diagnosis and therapy. *Endocrine Rev*. 2003;24(4):389–427.
- Halmos G, Schally AV, Sun BD, et al. High expression of somatostatin receptors and messenger ribonucleic acid for its receptor subtypes in organ-confined and locally advanced human prostate cancers. *J Clin Endocrinol Metabolism*. 2000;85(7):2564–2571.
- Plonowski A, Schally AV, Nagy A, et al. Inhibition of PC-3 human androgen-independent prostate cancer and its metastases by cytotoxic somatostatin analogue AN-238. *Cancer Res*. 1999;59(8):1947–1953.
- Barcena DG, Schally AV, Buenfil MV, et al. Response of patients with advanced prostatic cancer to administration of somatostatin analog RC-160 (vapeptide) at the time of relapse. *Prostate*. 2003;56(3):183–191.

## Funding

This work was supported by the National Cancer Institute (U01-CA98233-07, U54CA155626, U01-CA98710-06, U01-CA98216-06, U01-CA98758-07, CA141298), National Cancer Institute of Canada (019894), Hellenic Health Foundation, and Associazione Italiana per la Ricerca sul Cancro-AIRC-Milan.

## Notes

The study sponsors had no role in the design of the study; the collection, analysis, and interpretation of the data; the writing of the manuscript; and the decision to submit the manuscript for publication.

**Affiliations of authors:** Department of Epidemiology (YC, SL, MS, RT, DT, PK, JM), Department of Nutrition (YC, ELG, MS, WCW), and Department of Biostatistics (BR, XL), Harvard School of Public Health, Boston, MA; Channing Division of Network Medicine, Brigham Women's Hospital and Harvard Medical School, Boston, MA (YC, ELG, RT, MS, JM); Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA (FS, CAH, BH, DOS); Epidemiology Research Program, American Cancer Society, Atlanta, GA (VLS, WRD, SMG); Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD (DA, SIB, SJC, MY); Department of Epidemiology, German Institute of Human Nutrition Potsdam-Rehbruecke, Nuthetal, Germany (HB); National Institute for Public Health and the

Environment, Bilthoven, The Netherlands (HBB-d-M); Department of Gastroenterology and Hepatology, University Medical Centre, Utrecht, The Netherlands (HBB-d-M); Genomic Epidemiology Group (C055), German Cancer Research Center–Deutsches Krebsforschungszentrum, Heidelberg, Germany (FC); Public Health Division of Gipuzkoa, Epidemiology Unit, Basque regional Health Department and Biodonostia, Gipuzkoa, Spain (SC); Core Genotyping Facility Frederick National Laboratory for Cancer Research, Gaithersburg, MD (SJC, MY); Division of Preventive Medicine and Division of Aging, Department of Medicine, Brigham and Women's Hospital, Boston, MA (JMG); Massachusetts Veterans Epidemiology and Research Information Center and Geriatric Research, Education, and Clinical Center, Boston Veterans Affairs Healthcare System, Boston, MA (JMG); International Agency for Research on Cancer, Lyon, France (MJ); Department of Biobank Research, Umeå University, Sweden (MJ); University of Hawaii Cancer Center, Honolulu, HI (LLM); Molecular and Nutritional Epidemiology Unit Cancer Research and Prevention Institute–ISPO, Florence, Italy (DP); Genomics of Common Disease, School of Public Health, Imperial College London, London, UK (AS); Cancer Epidemiology Unit, University of Oxford, Oxford, UK (RCT); Bureau of Epidemiologic Research, Academy of Athens, Athens, Greece (DT); Hellenic Health Foundation, Athens, Greece (DT); Cancer Prevention Institute of California, Stanford Cancer Institute, CA (AWH); Departments of Medicine and Oncology, McGill University, Montreal, Canada (MP).

# Whole Milk Intake Is Associated with Prostate Cancer-Specific Mortality among U.S. Male Physicians<sup>1–4</sup>

Yan Song,<sup>5,6</sup> Jorge E. Chavarro,<sup>8,10,11</sup> Yin Cao,<sup>8,11</sup> Weiliang Qiu,<sup>8</sup> Lorelei Mucci,<sup>8,11</sup> Howard D. Sesso,<sup>9,11</sup> Meir J. Stampfer,<sup>8,10,11</sup> Edward Giovannucci,<sup>8,10,11</sup> Michael Pollak,<sup>12</sup> Simin Liu,<sup>5–7</sup> and Jing Ma<sup>8,11\*</sup>

<sup>5</sup>Department of Epidemiology and Program on Genomics and Nutrition, Fielding School of Public Health, <sup>6</sup>Center for Metabolic Disease Prevention, and <sup>7</sup>Departments of Medicine and Obstetrics and Gynecology, David Geffen School of Medicine, University of California, Los Angeles, CA; <sup>8</sup>Channing Division of Network Medicine, and <sup>9</sup>Divisions of Preventive Medicine and Aging, Brigham and Women's Hospital and Harvard Medical School; Boston, MA; <sup>10</sup>Department of Nutrition, and <sup>11</sup>Department of Epidemiology, Harvard School of Public Health, Boston, MA; and <sup>12</sup>Cancer Prevention Research Unit, Departments of Medicine and Oncology, Lady Davis Research Institute of the Jewish General Hospital and McGill University, Montreal, Quebec, Canada

## Abstract

Previous studies have associated higher milk intake with greater prostate cancer (PCa) incidence, but little data are available concerning milk types and the relation between milk intake and risk of fatal PCa. We investigated the association between intake of dairy products and the incidence and survival of PCa during a 28-y follow-up. We conducted a cohort study in the Physicians' Health Study ( $n = 21,660$ ) and a survival analysis among the incident PCa cases ( $n = 2806$ ). Information on dairy product consumption was collected at baseline. PCa cases and deaths ( $n = 305$ ) were confirmed during follow-up. The intake of total dairy products was associated with increased PCa incidence [HR = 1.12 (95% CI: 0.93, 1.35); >2.5 servings/d vs. #0.5 servings/d]. Skim/low-fat milk intake was positively associated with risk of low-grade, early stage, and screen-detected cancers, whereas whole milk intake was associated only with fatal PCa [HR = 1.49 (95% CI: 0.97, 2.28); \$237 mL/d (1 serving/d) vs. rarely consumed]. In the survival analysis, whole milk intake remained associated with risk of progression to fatal disease after diagnosis [HR = 2.17 (95% CI: 1.34, 3.51)]. In this prospective cohort, higher intake of skim/low-fat milk was associated with a greater risk of nonaggressive PCa. Most importantly, only whole milk was consistently associated with higher incidence of fatal PCa in the entire cohort and higher PCa-specific mortality among cases. These findings add further evidence to suggest the potential role of dairy products in the development and prognosis of PCa. *J. Nutr.* 143: 189–196, 2013.

## Introduction

Prostate cancer (PCa)<sup>13</sup> is one of the most common cancers among elderly men (1,2). Dairy product intake has been associated with higher risk of PCa in many (3–9) but not all (10–12) studies. In the Physicians' Health Study (PHS), we previously reported that higher intake of dairy products and

dairy-derived calcium were associated with a greater risk of developing incident PCa, based on 11 y of follow-up (9). Compared with men consuming #0.5 servings/d of dairy products, those consuming >2.5 servings/d had a 34% increase in risk of developing PCa (95% CI: 4%, 71%). In 2 meta-analyses of the relation between dairy product intake and PCa incidence, one showed a significant positive association (13), whereas the other reported an overall null association (14). Part of the reason for this inconsistency could be that most cohort studies (including our previous report in the PHS) and the 2 meta-analyses did not separately evaluate whole milk and skim/low-fat milk. In addition, most studies did not consider advanced disease or PCa-specific death as a major outcome, partly due to the variable duration of follow-up.

In the present study, we assessed the relation between intakes of types of dairy products and PCa risk, with a special emphasis on cases that were high grade and in advanced stages at diagnosis as well as the occurrence of fatal PCa during a 28-y follow-up.

<sup>1</sup> Supported by the NIH grants, CA42182, CA141298, CA104180, CA131945, CA097193, the Transdisciplinary Research in Energetics and Cancer Center U54CA155626, and the National Cancer Institute of Canada grant (019894). Y.S. is also supported by the UCLA Burroughs Wellcome Fund Inter-school Training Program in Metabolic Diseases (BWF-IT-MD).

<sup>2</sup> Author disclosures: Y. Song, J. E. Chavarro, Y. Cao, W. Qiu, L. Mucci, H. D. Sesso, M. J. Stampfer, E. Giovannucci, M. Pollak, S. Liu, and J. Ma, no conflicts of interest.

<sup>3</sup> Parts of this study were presented in abstract form at the 2012 UICC World Cancer Congress on August 28, 2012 in Montreal, Canada.

<sup>4</sup> This trial was registered at [www.clinicaltrials.gov](http://www.clinicaltrials.gov) as NCT00000500.

<sup>13</sup> Abbreviations used: FDR, false-discovery rate; PCa, prostate cancer; PHS, Physicians' Health Study; PSA, prostate-specific antigen.

\* To whom correspondence should be addressed. E-mail: [jing.ma@channing.harvard.edu](mailto:jing.ma@channing.harvard.edu).

## Participants and Methods

**Study population.** The PHS was a randomized, blinded, and placebo-controlled trial of aspirin and *b*-carotene in the prevention of heart disease and cancer among 22,071 U.S. male physicians aged 40–84 y in 1982 (15,16). At enrollment, participants provided information in the enrollment questionnaires on medical history and several lifestyle factors. All the physicians who were eligible and willing to participate were enrolled in a run-in phase. After 18 wk, participants were sent a questionnaire asking about their health status, side effects of treatment, compliance, and willingness to continue in the trial. Follow-up questionnaires were mailed at 6 and 12 mo after randomization and annually thereafter. Participants were asked to report newly diagnosed diseases, including PCa. For this study, we limited the study population to men who returned the run-in questionnaires with relevant abbreviated dietary information. To reduce the potential for undiagnosed PCa to influence diet and to utilize the dietary data collected on the 12-mo questionnaire, we excluded PCa cases diagnosed during the first year in the study, men with BMI <18.5 kg/m<sup>2</sup> at baseline, and men without baseline BMI information. These exclusions resulted in a study population of 21,660 men for analysis. The study design and methods used in this investigation were reviewed and approved by the Institutional Review Board of Partners Healthcare.

**Dietary assessment.** The run-in and 12-mo questionnaires in the PHS included abbreviated FFQs. The run-in questionnaire asked about the consumption of whole milk, skim/low-fat milk, and cold breakfast cereal (categories: \$2 servings/d, daily, 5–6 servings/wk, 2–4 servings/wk, 1 serving/wk, 1–3 servings/mo, rarely/never) in the past year. The 12-mo questionnaire asked about the intake during the previous year of hard cheese (e.g., American, Cheddar) and ice cream. We considered these 5 foods to be the main contributors to dairy product intake and combined those responses by servings to estimate total daily dairy product intake (9). Because the potential effects of dairy calcium on PCa risk were of interest, we also calculated total dairy calcium intake from each dairy product. Calcium content was obtained from the nutrient composition database of the USDA (17). The calcium content per serving (as weights in the total calcium consumption) is as follows: whole milk (1 serving = 237 mL), 276 mg; skim/low-fat milk (1 serving = 237 mL), 299 mg; ice cream (1 serving = 214 g, as in vanilla savor), 169 mg; and hard cheese (1 serving = 28 g, as an average of American cheese and Cheddar cheese), 173 mg. Two questions about red meat intake were also included in the 12-mo questionnaire, which asked about the intake of beef, pork or lamb as a sandwich or mixed dish (hamburger, stew, casserole, lasagna, etc.) and those as a main dish (steak, roast, ham, etc.). Daily intake of red meat was calculated as the sum of the servings (1 serving = 227 g) for each of these 2 items.

**Ascertainment of PCa outcomes.** For the PCa incidence analyses, men were followed from the date when the 12-mo questionnaire was returned until the date of PCa diagnosis, date of death, or the end of follow-up (March 9, 2010), whichever came first. For the PCa-specific analyses, men were followed from the date of PCa diagnosis until the date of death from PCa, date of death from other causes, or March 9, 2010, whichever came first. We learned of deaths in the cohort through notification by family members and postal authorities and through periodic systematic searches of the National Death Index. Cause of death was determined by an endpoint committee of 3 physicians based on all available information, including medical records and death certificates. Follow-up for mortality was at least 97.7% complete and for morbidity, 95.3% (18).

Whenever a participant reported a new diagnosis of PCa, we requested hospital records and pathology reports to confirm the diagnosis and determine tumor stage, grade, and other clinical characteristics at diagnosis. Histological grade was recorded following the Gleason scoring system from the pathology reports. Low-grade tumors were defined as Gleason #7 and high-grade was defined as Gleason >7. Clinical stage was determined using the TNM staging system. Tumors of stage T3 or higher (T3/T4/N1/M1) were categorized as advanced-stage tumors and tumors of stage T1 or T2 were defined as early-stage tumors. Cases without pathological staging were classified as undetermined stage unless there

was clinical evidence of distant metastases. Because prostate-specific antigen (PSA) screening has dramatically changed the clinical presentation of the cancer, we also categorized the cases into 3 groups: pre-PSA era cases (diagnosed before 1990), post-PSA era cases (diagnosed 1990 or thereafter) who presented with prostatic or metastatic symptoms, and post-PSA era cases detected by PSA or digital rectal examination screening.

**Statistical analyses.** To examine the association of dairy products and calcium consumption with PCa risk, we used Cox proportional hazards regression models to calculate the HR and 95% CI, with the lowest intake category as the reference group. We categorized the intake of each dairy food into 4 groups (rarely, #1 serving/wk, 2–6 servings/wk, and \$1 serving/d). Calcium intake from dairy products was categorized into 5 groups by quintiles. Tests for linear trend were performed using the median intake values in each category as a continuous variable. Beyond age-adjusted models, multivariable models additionally included terms for baseline (time when 12-mo questionnaire was returned) cigarette smoking (never, past, or current smoker), vigorous exercise (exercise vigorously to sweat more than twice per week or not), alcohol intake (drink alcoholic beverages every day or not), race (Caucasian or non-Caucasian), BMI (<25.0, 25.0–29.9, or \$30.0 kg/m<sup>2</sup>), diabetes status (yes or no), red meat consumption (servings/week), and assignment in the original trial (active treatment or placebo for aspirin and *b*-carotene). In addition, the models for whole milk and skim/low-fat milk were mutually adjusted for each other.

The abbreviated FFQs in the PHS were not comprehensive; thus, we were unable to calculate and adjust for total energy intake directly. To minimize the potential confounding due to total energy intake, we calculated total energy intake using only the food items that were recorded in the run-in and 12-mo questionnaires. These food items included 13 types of fruits and vegetables, 5 types of dairy foods investigated in this study, eggs, chicken, beef, 4 types of fish and seafood, cookies, chips, nuts, and fried foods. Under similar situations, previous studies used food scores by summing up servings of all recorded food items (9,19). In this study, we weighted the servings of recorded food items with total calorie per serving of each individual item to better emulate total energy intake calculated from comprehensive FFQs.

Separate multivariable models for PCa incidence were fit for subgroups of cancer according to Gleason grade, clinical stage, and disease presentation at diagnosis, and disease fatality during follow-up. We then modeled the relation between dairy product and PCa-specific mortality among cases using the Cox proportional hazard regression model. Besides the age- and multivariable-adjusted model [including the same set of covariates as in the incidence model and stage of tumor (T3/T4/N1/M1 or T1/T2) and Gleason score (>7 or #7)], we further stratified the analyses by disease presentation at diagnosis (pre-PSA era presented, post-PSA era presented by symptom, and post-PSA era presented by screening). To account for potential false positives due to multiple comparisons, we calculated the false-discovery rate (FDR) by incorporating all P values from multiple tests performed for the linear trends. The FDR statistics were obtained for each P value, and FDR statistics with  $q < 0.05$  were considered significant (20). All analyses were performed in SAS version 9.3 (SAS Institute). All P values are 2-sided.

## Results

We confirmed 2806 incident cases of PCa diagnosed among 21,660 men in 470,612 person-years through 2010. The baseline characteristics of the study population by categories of dairy product intake are presented in Table 1. Men who consumed more dairy products tended to be older, smoked less, drank less alcohol, exercised more, and were more likely to be Caucasian and diabetic. When stratified by type of milk, the data showed that men who consumed more skim/low-fat milk tended to smoke less, drink less alcohol, and exercise more and were more likely to be Caucasian, whereas men who consumed more

TABLE 1 Baseline characteristics by category of baseline dairy product intake in the PHS ( $n = 21,660$ )<sup>1</sup>

	Dairy product intake, <sup>2</sup> servings					P	Whole milk, servings		P	Skim/low-fat milk, servings		P
	#0.5/d (n = 3446)	.0.5–1.0/d (n = 3878)	.1.0–1.5/d (n = 4527)	.1.5–2.5/d (n = 6390)	.2.5/d (n = 3302)		#1/wk (n = 16,618)	\$2/wk (n = 4207)		#1/wk (n = 11,834)	\$2/wk (n = 9186)	
Age, y	52.3	52.7	53.4	54.3	55.1	.001	53.0	55.5	.001	53.4	53.6	.12
BMI, %						.001			.08			.004
Normal weight	56	55	56	60	61		58	56		57	59	
Overweight	40	41	40	36	35		38	39		39	37	
Obese	4	4	4	4	4		4	4		4	4	
Caucasian, %	85	91	93	95	96	.001	93	89	.001	90	95	.001
Diabetes, %	1.6	1.4	1.5	2.4	2.6	.001	1.7	2.6	.001	1.6	2.3	.001
Smoking, %						.001			.001			.001
Never	46	47	49	53	56		51	48		48	54	
Former	41	41	40	39	35		40	37		41	37	
Current	13	12	11	8	9		9	15		12	8	
Frequent drinker, <sup>3</sup> %	29	25	24	24	20	.001	24	24	.97	26	21	.001
Vigorous exercise, <sup>4</sup> %	48	52	54	56	59	.001	55	50	.001	51	58	.001
Red meat intake, <sup>5</sup> servings/wk	0.6	0.7	0.7	0.7	0.8	.001	0.7	0.8	.001	0.7	0.7	.001

<sup>1</sup> Values are percentage or mean  $\pm$  SE. PHS, Physicians' Health Study.

<sup>2</sup> Based on the consumption of 5 major dairy foods (whole milk, skim/low-fat milk, hard cheese, ice cream, and cold breakfast cereal) assessed from 1982 to 1984. One serving of whole milk, skim/low-fat milk, or cold breakfast cereal = 237 mL; 1 serving of ice cream = 214 g; 1 serving of hard cheese = 28 g.

<sup>3</sup> Frequent drinker was defined as someone who drinks alcoholic beverages every day.

<sup>4</sup> Vigorous exercise was defined as to exercise vigorously to a sweat more than twice per week.

<sup>5</sup> 1 serving of red meat = 227 g.

whole milk tended to be current smokers, exercise less, and less likely to be Caucasian.

Total dairy food intake was marginally associated with overall PCa risk. In multivariable-adjusted analyses, men in the highest category of total dairy foods had a 12% (95% CI: 27%, 35%) higher risk to develop PCa than men in the lowest intake category (P-trend = 0.06) (Table 2). For individual dairy foods, skim/low-fat milk had the strongest association with PCa incidence: the multivariable-adjusted HR was 1.19 (95% CI: 1.06, 1.33; P-trend = 0.001), comparing the highest [237 mL/d (1 serving/d)] with the lowest (rarely consumed) intake category. In contrast, whole milk, hard cheese, ice cream, and cold breakfast cereal intakes were not significantly associated with overall risk of PCa incidence. Calcium from dairy foods was marginally associated with PCa incidence (P-trend = 0.07).

We next examined the association of total dairy products, whole milk, and skim/low-fat milk with special attention to cancer subtypes and the timing of diagnosis (i.e., 1982–1989, pre-PSA era vs. 1990–2010, post-PSA era) (Table 3). We found that higher intake of skim/low-fat milk was mainly associated with a higher risk of low-grade, early-stage, and screen-detected disease; comparing the highest with the lowest intake category, the HR were 1.20 for low-grade cases (95% CI: 1.06, 1.37), 1.19 for early-stage cases (95% CI: 1.04, 1.35), and 1.21 for post-PSA era cases detected by screening (95% CI: 1.02, 1.43) (P-trend # 0.01 for all the subgroup analyses). In contrast, for risk of fatal PCa, whole milk was the only dairy food that had a positive association [HR = 1.49 (95% CI: 0.97, 2.28); P-trend = 0.01]. This association was independent of age, cigarette smoking status, BMI, alcohol intake, vigorous physical activity, diabetes status, red meat consumption, and total energy intake from recorded food items.

Finally, among all the PCa cases, we conducted a survival analysis to evaluate the associations of prediagnostic dairy food intake with risk of progression to fatal PCa after initial diagnosis and found that whole milk was the only dairy food that was significantly associated with an increased risk of PCa-specific

mortality (Table 4). Compared with nondrinkers of whole milk, the multivariable-adjusted HR was 2.17 (95% CI: 1.34, 3.51; P-trend < 0.001) for those who consumed 237 mL/d (1 serving/d). A stratified analysis on age at diagnosis showed that high intake of whole milk was significantly associated with risk of progression to fatal PCa in both old and young age groups, except that there tended to be a J-shaped relation in the older group (data not shown). In a stratified analysis on the presentation of disease, we found that, among post-PSA era cases presented by screening, whole milk intake was associated with PCa deaths, although the  $q$  value was not significant [HR = 1.82 (95% CI: 0.69, 4.84); P-trend = 0.07]. The associations with skim/low-fat milk, however, were not significant in any of the strata by PSA era and screening.

## Discussion

In this study, we confirmed and extended our previous findings that total dairy product intake and calcium from dairy foods were positively associated with overall risk of PCa. Admittedly, the dairy variables in our study did not capture all dairy product intake (did not include information on intakes of yogurt, cream, butter, etc.). However, according to data from the NHANES, milk and cheese intakes can account for 98% of total dairy product intake (21). Thus, our data on available dairy food items sufficiently represented the total dairy product intake in our population. The magnitude of the overall association between total dairy product intake and the risk of incident PCa [HR = 1.12 (95% CI: 0.93, 1.35)] in this study, however, was weaker than in our previous report [RR = 1.34 (95% CI: 1.04, 1.71)]. Because the current analysis had a much larger sample size (2806 cases vs. 1012 cases) and an additional 15 y of follow-up, these allowed us to specifically evaluate subtypes of dairy products and by subtypes of PCa, cancer diagnosed before vs. in the PSA era, mode of diagnosis, and cancer-specific mortality (9). We found that skim/low-fat milk intake were

TABLE 2 HR estimates for PCa by intake of dairy product and dairy calcium in the PHS ( $n = 21,660$ )<sup>1</sup>

	Category 1	Category 2	Category 3	Category 4	Category 5	P-trend <sup>2</sup>
All dairy food <sup>3</sup>						
Cases/person-years	388/76,216	446/86,740	586/98,871	910/137,667	458/69,738	
Age-adjusted	1.00	1.00 (0.88, 1.15)	1.11 (0.98, 1.26)	1.19 (1.06, 1.34)	1.15 (1.00, 1.31)	0.003 <sup>4</sup>
Multivariable-adjusted <sup>5</sup>	1.00	0.96 (0.83, 1.11)	1.07 (0.93, 1.23)	1.15 (0.99, 1.32)	1.12 (0.93, 1.35)	0.06
Whole milk <sup>6</sup>						
Cases/person-years	1674/279,675	504/86,554	273/47,723	244/39,924		
Age-adjusted	1.00	0.97 (0.88, 1.08)	0.89 (0.78, 1.01)	0.89 (0.78, 1.02)		0.04
Multivariable-adjusted <sup>5</sup>	1.00	1.02 (0.92, 1.13)	0.93 (0.81, 1.07)	0.95 (0.81, 1.10)		0.32
Skim/low-fat milk <sup>6</sup>						
Cases/person-years	895/160,367	531/98,250	579/94,591	724/104,959		
Age-adjusted	1.00	1.05 (0.94, 1.17)	1.17 (1.05, 1.29)	1.21 (1.10, 1.34)		.0001 <sup>4</sup>
Multivariable-adjusted <sup>5</sup>	1.00	1.02 (0.91, 1.14)	1.12 (1.00, 1.25)	1.19 (1.06, 1.33)		0.001 <sup>4</sup>
Hard cheese <sup>6</sup>						
Cases/person-years	197/35,560	1207/208,462	1175/190,531	178/28,270		
Age-adjusted	1.00	1.05 (0.90, 1.22)	1.12 (0.96, 1.30)	1.10 (0.90, 1.35)		0.14
Multivariable-adjusted <sup>5</sup>	1.00	1.01 (0.87, 1.18)	1.07 (0.91, 1.25)	1.05 (0.85, 1.30)		0.32
Ice cream <sup>6</sup>						
Cases/person-years	455/75,120	1415/251,406	805/124,783	84/12,177		
Age-adjusted	1.00	0.96 (0.86, 1.06)	1.06 (0.95, 1.19)	1.05 (0.83, 1.32)		0.06
Multivariable-adjusted <sup>5</sup>	1.00	0.95 (0.85, 1.06)	1.02 (0.90, 1.15)	1.03 (0.80, 1.32)		0.26
Cold breakfast cereal <sup>6</sup>						
Cases/person-years	743/131,310	654/120,759	678/112,540	679/98,469		
Age-adjusted	1.00	0.96 (0.86, 1.06)	1.02 (0.92, 1.13)	1.11 (1.00, 1.23)		0.01 <sup>4</sup>
Multivariable-adjusted <sup>5</sup>	1.00	0.95 (0.85, 1.06)	1.00 (0.88, 1.12)	1.06 (0.93, 1.22)		0.17
Calcium from dairy food <sup>7</sup>						
Cases/person-years	487/95,147	516/95,489	578/93,334	598/92,688	609/91,575	
Age-adjusted	1.00	1.04 (0.92, 1.18)	1.15 (1.02, 1.30)	1.16 (1.03, 1.31)	1.17 (1.03, 1.31)	0.004 <sup>4</sup>
Multivariable-adjusted <sup>5</sup>	1.00	1.01 (0.89, 1.15)	1.12 (0.98, 1.28)	1.12 (0.97, 1.30)	1.14 (0.97, 1.34)	0.07

<sup>1</sup> Values are HR (95% CI). FDR, false-discovery rate; PCa, prostate cancer; PHS, Physicians' Health Study.

<sup>2</sup> Calculated in a separate regression model with the median intake levels in each category as a continuous variable.

<sup>3</sup> Based on the consumption of 5 major dairy foods (whole milk, skim/low-fat milk, hard cheese, ice cream, and cold breakfast cereal) assessed from 1982 to 1984. The 5 intake level groups are: #0.5 servings/d, .05–1.0 serving/d, .10–1.5 servings/d, .15–2.5 servings/d, and .25 servings/d. One serving of whole milk, skim/low-fat milk, or cold breakfast cereal = 237 mL; 1 serving of ice cream = 214 g; 1 serving of hard cheese = 28 g.

<sup>4</sup> FDR = 0.05.

<sup>5</sup> Adjusted for baseline measures of age (y), cigarette smoking (never, past, current), vigorous exercise (exercise vigorously to a sweat more than twice per week or not), alcohol intake (drink alcoholic beverages every day or not), race (Caucasian, non-Caucasian), BMI (normal weight, overweight, obese), baseline diabetes status (yes, no), red meat consumption (servings/wk), total energy intake from recorded food items (kcal), assignment in the original aspirin trial (treatment, placebo), and assignment in the original *b*-carotene trial (treatment, placebo). In addition, the models for whole milk and skim/low-fat milk were mutually adjusted for each other (rarely, #1 serving/wk, 2–6 servings/wk, and \$1 serving/d).

<sup>6</sup> The 4 intake level groups were: rarely, #1 serving/wk, 2–6 servings/wk, and \$1 serving/d.

<sup>7</sup> The 5 intake level groups were categorized according to quintiles.

related to a higher risk of nonaggressive disease (low-grade, early-stage, and screen-detected cases), whereas whole milk intake was associated with a higher risk of fatal PCa and, among all the cases, with a higher risk of progression to fatal PCa.

The positive association between dairy product intake and PCa has been reported in several studies, including the European Prospective Investigation into Cancer and Nutrition (22) and studies from Canada (23) and Japan (4). These data raised concerns regarding whether dairy should be recommended as part of a healthy diet for aging men (24,25). However, the results of 2 meta-analyses of the relation between dairy product intake and PCa provided conflicting conclusions: one showed a significant positive association (13) and the other (supported by the National Dairy Council) showed an overall null association (14). Part of the reason for this inconsistency could be a lack of detailed data for the effect of whole compared with skim/low-fat milk and their impact on high-risk disease or PCa-specific death.

Our finding that the strongest association with total dairy products was in the pre-PSA era was consistent with findings of Rodriguez et al. (26). We observed a significant positive association of skim/low-fat milk with overall PCa risk. These results are consistent with previous studies (6,27). Few studies specifically evaluated high-risk PCa. Park et al. (28) observed that skim milk, but not other dairy foods, was associated with a nonsignificantly increased risk of advanced PCa. The null effect of whole milk on overall PCa risk is likely due to the fact that the whole milk drinkers accounted for only a small portion of all milk drinkers. Thus, the associations of whole milk with the nonfatal cases, if any, were not large enough to be detected with a limited number of cases, which may have driven the overall effect.

The commonly accepted risk factors for incident PCa are older age, a family history of PCa, and being African American (29). However, there is no consensus about risk factors for fatal PCa beyond clinical characteristics such as PSA at diagnosis, Gleason grade, and clinical stage. Identifying modifiable risk

TABLE 3 Multivariable-adjusted HR estimates for categories of PCa cases by intake of dairy product in the PHS ( $n = 21,660$ )<sup>1,2</sup>

Selected case <sup>3</sup>	Category 1	Category 2	Category 3	Category 4	Category 5	P-trend <sup>4</sup>
<b>Dairy product<sup>5</sup></b>						
High grade	1.00	1.04 (0.69, 1.58)	0.77 (0.50, 1.20)	1.09 (0.71, 1.68)	1.04 (0.60, 1.80)	0.64
Low grade	1.00	0.95 (0.81, 1.12)	1.11 (0.95, 1.30)	1.13 (0.95, 1.33)	1.13 (0.91, 1.39)	0.12
Advanced	1.00	0.92 (0.59, 1.46)	0.79 (0.50, 1.27)	0.92 (0.57, 1.48)	0.68 (0.36, 1.27)	0.35
Localized	1.00	0.94 (0.80, 1.11)	1.09 (0.93, 1.29)	1.11 (0.94, 1.32)	1.13 (0.91, 1.39)	0.13
Fatal	1.00	1.19 (0.68, 2.06)	1.81 (1.08, 3.02)	2.14 (1.26, 3.64)	1.73 (0.90, 3.35)	0.05
Pre-PSA	1.00	1.70 (0.95, 3.05)	1.77 (1.00, 3.13)	1.82 (1.01, 3.27)	2.12 (1.07, 4.19)	0.10
Post-PSA (symptom)	1.00	1.44 (0.78, 2.68)	1.25 (0.66, 2.34)	1.83 (0.99, 3.40)	1.61 (0.76, 3.40)	0.19
Post-PSA (screening)	1.00	0.83 (0.67, 1.03)	1.10 (0.90, 1.34)	1.04 (0.84, 1.28)	0.99 (0.75, 1.30)	0.64
<b>Whole milk<sup>6</sup></b>						
High grade	1.00	0.69 (0.48, 1.00)	1.29 (0.91, 1.84)	0.78 (0.49, 1.25)		0.81
Low grade	1.00	1.09 (0.97, 1.23)	0.86 (0.73, 1.01)	0.91 (0.76, 1.10)		0.10
Advanced	1.00	0.89 (0.61, 1.29)	1.04 (0.68, 1.61)	0.83 (0.49, 1.41)		0.63
Localized	1.00	1.06 (0.94, 1.19)	0.87 (0.74, 1.03)	0.89 (0.74, 1.07)		0.08
Fatal	1.00	0.89 (0.60, 1.31)	1.77 (1.23, 2.54)	1.49 (0.97, 2.28)		0.01 <sup>7</sup>
Pre-PSA	1.00	1.29 (0.89, 1.86)	1.51 (1.00, 2.27)	1.35 (0.85, 2.15)		0.15
Post-PSA (symptom)	1.00	1.22 (0.80, 1.86)	1.19 (0.71, 1.99)	1.29 (0.76, 2.21)		0.38
Post-PSA (screening)	1.00	1.00 (0.86, 1.17)	0.74 (0.59, 0.93)	0.73 (0.57, 0.94)		0.002 <sup>7</sup>
<b>Skim/low-fat milk<sup>6</sup></b>						
High grade	1.00	1.11 (0.79, 1.56)	1.07 (0.76, 1.51)	1.19 (0.85, 1.67)		0.39
Low grade	1.00	0.99 (0.87, 1.13)	1.18 (1.04, 1.35)	1.20 (1.06, 1.37)		0.001 <sup>7</sup>
Advanced	1.00	0.94 (0.64, 1.38)	1.02 (0.69, 1.49)	0.99 (0.67, 1.45)		0.96
Localized	1.00	0.99 (0.86, 1.13)	1.18 (1.04, 1.35)	1.19 (1.04, 1.35)		0.004 <sup>7</sup>
Fatal	1.00	1.04 (0.72, 1.51)	1.01 (0.69, 1.47)	1.04 (0.71, 1.51)		0.91
Pre-PSA	1.00	1.38 (0.92, 2.07)	1.69 (1.15, 2.48)	1.43 (0.97, 2.12)		0.11
Post-PSA (symptom)	1.00	0.84 (0.52, 1.35)	1.02 (0.64, 1.62)	1.22 (0.79, 1.88)		0.23
Post-PSA (screening)	1.00	0.98 (0.83, 1.17)	1.20 (1.01, 1.42)	1.21 (1.02, 1.43)		0.01 <sup>7</sup>

<sup>1</sup> Values are HR (95% CI). FDR, false-discovery rate; PCa, prostate cancer; PHS, Physicians' Health Study; PSA, prostate-specific antigen.

<sup>2</sup> Adjusted for baseline measures of age (y), cigarette smoking (never, past, current), vigorous exercise (exercise vigorously to a sweat more than twice per week or not), alcohol intake (drink alcoholic beverages every day or not), race (Caucasian, non-Caucasian), BMI (normal weight, overweight, obese), baseline diabetes status (yes, no), and red meat consumption (servings/wk), total energy intake from recorded food items (kcal), assignment in the original aspirin trial (treatment, placebo), and assignment in the original *b*-carotene trial (treatment, placebo). In addition, the models for whole milk and skim/low-fat milk were mutually adjusted for each other (rarely, #1 serving/wk, 2–6 servings/wk, and \$1 serving/d).

<sup>3</sup> High grade ( $n = 317$ ): Gleason  $\geq 7$ ; low grade ( $n = 2105$ ): Gleason  $\leq 6$ ; advanced ( $n = 272$ ): T3/T4/N1/M1; localized ( $n = 2016$ ): T1/T2; fatal ( $n = 305$ ): died of PCa; pre-PSA era ( $n = 274$ ): diagnosed before 1990; post-PSA era: diagnosed after 1990; presented by symptom ( $n = 192$ ): presented by prostate-related symptoms or metastases; presented by screening ( $n = 1233$ ): presented by PSA test screening or digital rectal examination;

<sup>4</sup> Calculated in a separate regression model with the median intake in each category as a continuous variable.

<sup>5</sup> Based on baseline consumption of 5 major dairy foods (whole milk, skim/low-fat milk, hard cheese, ice cream, and cold breakfast cereal). The 5 intake level groups are: #0.5 servings/d, .05–1.0 serving/d, .10–1.5 servings/d, .15–2.5 servings/d, and .25 servings/d. One serving of whole milk, skim/low-fat milk, or cold breakfast cereal = 237 mL; 1 serving of ice cream = 214 g; 1 serving of hard cheese = 28 g.

<sup>6</sup> The 4 intake level groups were: rarely, #1 serving/wk, 2–6 servings/wk, and \$1 serving/d.

<sup>7</sup> FDR  $\leq 0.05$ .

factors for fatal PCa is critical, because widespread PSA testing in the US is likely to detect and overtreat a large number of men with indolent cancer (30). A major challenge in PCa research is distinguishing risk factors for aggressive PCa from indolent disease to reduce overtreatment. Our results showed that higher intakes of whole-fat milk predispose men to a higher risk of developing fatal PCa and, once they had the cancer, a higher risk of progression to fatal disease. This association was unlikely confounded by skim/low-fat milk according to our analysis.

Given that dairy product intakes were assessed years before cancer diagnosis, our findings need to be further confirmed by cohorts with more detailed dietary information, especially dietary intakes at or around the time of the cancer diagnosis. In the Health Professionals Follow-up Study cohort, Chan et al. (31) found that men in the highest compared with the lowest quartile of milk consumption after diagnosis had a

nonsignifi elevated risk of fatal PCa [HR = 1.30 (95% CI: 0.93, 1.83)], but this study did not examine specific types of dairy food. Another explanation of the association between whole milk intake and fatal PCa risk is also possible: it is likely that men who drink more whole milk are less likely to be screened and therefore are diagnosed at a later stage and are at a higher risk for fatal disease. In the survival analysis, we adjusted for Gleason score and stage of tumor at diagnoses. The association remained significant after the adjustment, which supports that the association was not due to confounding by screening. However, further data on PSA screening intensity are needed to justify or refute this explanation.

In our study, the average interval between dairy product intake assessment and PCa diagnosis was 14 y, yielding possible exposure misclassification. This is of particular concern for the analysis of PCa survival, because patients may have changed

TABLE 4 HR estimates of PCa death by prediagnostic intake of dairy product and dairy calcium in PCa cases in the PHS ( $n = 2806$ )<sup>1</sup>

	Category 1	Category 2	Category 3	Category 4	Category 5	P-trend <sup>2</sup>
All dairy food <sup>3</sup>						
Deaths/person-years	27/3601	45/4012	74/5222	115/8503	43/4416	
Age-adjusted	1.00	1.50 (0.93, 2.41)	1.83 (1.18, 2.85)	1.73 (1.14, 2.63)	1.22 (0.75, 1.97)	0.75
Multivariable-adjusted <sup>4,5</sup>	1.00	0.97 (0.53, 1.78)	2.23 (1.26, 3.92)	1.87 (1.04, 3.37)	1.71 (0.82, 3.58)	0.16
Pre-PSA	1.00	1.16 (0.33, 4.05)	2.20 (0.66, 7.29)	1.02 (0.29, 3.51)	2.15 (0.53, 8.79)	0.76
Post-PSA (screening)	1.00	0.74 (0.27, 2.03)	1.09 (0.45, 2.64)	1.24 (0.50, 3.06)	0.93 (0.26, 3.36)	0.80
Whole milk <sup>6</sup>						
Deaths/person-years	161/15,350	43/4860	49/2504	43/2092		
Age-adjusted	1.00	0.85 (0.60, 1.18)	1.81 (1.32, 2.49)	1.85 (1.32, 2.59)		.0001 <sup>7</sup>
Multivariable-adjusted <sup>4,5</sup>	1.00	0.73 (0.47, 1.13)	1.79 (1.15, 2.79)	2.17 (1.34, 3.51)		.0001 <sup>7</sup>
Pre-PSA	1.00	0.77 (0.36, 1.63)	0.68 (0.23, 1.98)	1.21 (0.45, 3.24)		0.67
Post-PSA (screening)	1.00	0.84 (0.40, 1.79)	2.26 (1.07, 4.78)	1.82 (0.69, 4.84)		0.07
Skim/low-fat milk <sup>6</sup>						
Deaths/person-years	115/8106	58/4856	53/5493	68/6789		
Age-adjusted	1.00	0.89 (0.65, 1.22)	0.71 (0.51, 0.98)	0.70 (0.52, 0.94)		0.02 <sup>7</sup>
Multivariable-adjusted <sup>4,5</sup>	1.00	1.01 (0.67, 1.52)	0.87 (0.56, 1.36)	1.02 (0.67, 1.56)		0.98
Pre-PSA	1.00	1.18 (0.53, 2.59)	0.61 (0.24, 1.56)	0.77 (0.34, 1.76)		0.38
Post-PSA (screening)	1.00	1.48 (0.71, 3.11)	1.34 (0.66, 2.73)	1.22 (0.54, 2.73)		0.79
Calcium from dairy food <sup>8</sup>						
Deaths/person-years	41/4412	52/4755	73/5174	74/5535	64/5879	
Age-adjusted	1.00	1.18 (0.79, 1.78)	1.45 (0.99, 2.13)	1.36 (0.93, 1.99)	1.11 (0.75, 1.64)	0.75
Multivariable-adjusted <sup>4,5</sup>	1.00	1.06 (0.63, 1.79)	1.70 (1.00, 2.89)	1.64 (0.94, 2.89)	1.71 (0.91, 3.21)	0.09
Pre-PSA	1.00	0.83 (0.33, 2.08)	1.30 (0.49, 3.48)	0.63 (0.23, 1.77)	1.30 (0.41, 4.19)	0.99
Post-PSA (screening)	1.00	1.09 (0.44, 2.69)	1.08 (0.45, 2.62)	1.63 (0.65, 4.10)	1.22 (0.41, 3.65)	0.59

<sup>1</sup> Values are HR (95% CI). FDR, false-discovery rate; PCa, prostate cancer; PHS, Physicians' Health Study; PSA, prostate-specific antigen.

<sup>2</sup> Calculated in a separate regression model with the median intake levels in each category as a continuous variable.

<sup>3</sup> Based on the consumption of 5 major dairy foods (whole milk, skim/low-fat milk, hard cheese, ice cream, and cold breakfast cereal) assessed from 1982 to 1984. The 5 intake level groups are: #0.5 servings/d, .05–1.0 serving/d, 1.0–1.5 servings/d, 1.5–2.5 servings/d, and 2.5 servings/d. One serving of whole milk, skim/low-fat milk, or cold breakfast cereal = 237 mL; 1 serving of ice cream = 214 g; 1 serving of hard cheese = 28 g.

<sup>4</sup> Adjusted for baseline measures of age at diagnosis (y), cigarette smoking (never, past, current), vigorous exercise (exercise vigorously to a sweat more than twice per week or not), alcohol intake (drink alcoholic beverages every day or not), race (Caucasian, non-Caucasian), BMI (normal weight, overweight, obese), baseline diabetes status (yes, no), red meat consumption (servings/wk), Gleason score (7, 8), stage of tumor (T3/T4/N1/M1, T1/T2), total energy intake from recorded food items (kcal), assignment in the original aspirin trial (treatment, placebo), and assignment in the original *b*-carotene trial (treatment, placebo). In addition, the models for whole milk and skim/low-fat milk were mutually adjusted for each other (rarely, #1 serving/wk, 2–6 servings/wk, and #1 serving/d).

<sup>5</sup> Pre-PSA era ( $n = 274$ ): diagnosed before 1990; post-PSA era: diagnosed after 1990; presented by symptom ( $n = 192$ ); presented by prostate-related symptoms or metastases (results not presented because of very low statistical power); presented by screening ( $n = 1233$ ); presented by PSA test screening or digital rectal examination.

<sup>6</sup> The 4 intake level groups were: rarely, #1 serving/wk, 2–6 servings/wk, and #1 serving/d.

<sup>7</sup> FDR = 0.05.

<sup>8</sup> The 5 intake level groups were categorized according to quintiles.

their diet after diagnosis. We evaluated correlations among nutrients between the 2000 and 2004 FFQs, comparing men diagnosed with PCa in that interval with those who remained free of PCa. We found that the correlations ranged between 0.5 and 0.7 for all nutrients assessed, including dairy products. There were no obvious trends in the absolute levels of intake between cases and non-cases. These observations suggest that men tended to keep their dietary habits after PCa diagnosis. One advantage of using prediagnostic dietary information is to avoid confounding by recall bias, change of diet due to disease severity or treatments, or other reasons. Recently, Pettersson et al. (32) found that in the Health Professionals Follow-up Study, post-diagnostic milk and dairy product intake was not significantly associated with increased risk of fatal PCa, whereas Torfadotir et al. (33) found that milk intake during adolescence, rather than in midlife or currently, was associated with advanced PCa. One possibility is that dairy product intake in earlier life may be more relevant to the progression and mortality of PCa in later life.

Several potential mechanisms could explain the observed associations of dairy food (primarily skim/low-fat milk) with overall PCa risk. First, skim/low-fat milk is the major source of dairy calcium and higher intake might lower intra-cellular 1,25-dihydroxycholecalciferol concentrations and induce prostate carcinogenesis (8,34–36). Second, the association could be mediated via phytanic acid, which may upregulate expression of  $\alpha$ -methylacyl-CoA racemase (37,38). The involvement of  $\alpha$ -methylacyl-CoA racemase in PCa is implicated by a recent observation (39). Third, the relation could be through the effect of phosphate. Newmark et al. (40) suggested that the high dietary phosphate content of dairy products might explain the risk of PCa induced by dairy products, because the plasma phosphate concentration can appreciably influence 1,25-dihydroxycholecalciferol concentrations. Fourth, the ability of dairy products to raise concentrations of insulin-like growth factor 1 have also been suggested as a possible explanation for the association (41–43). The association of whole milk with fatal PCa and

PCa-specific mortality may be via the effects of dairy fat (primarily saturated fat) or other factors (including obesity and hyperinsulinemia). Whole milk has an ;40 times higher content of saturated fat compared with skim milk and the difference of the saturated fat content between 237 mL of whole milk and skim milk is ;20% of its average daily intake (17). High-fat dairy has been positively correlated with higher C-peptide concentrations, which were positively related to risk of aggressive PCa (44).

In summary, the results from the present study confi a potential role of dairy products in PCa risk and survival. Skim/low-fat milk dairy products have been suggested as being benefi for several disease outcomes, including colorectal cancer; so future research is warranted to investigate the optimal intake of skim/low-fat dairy products. However, our results add further evidence to suggest that the intake of whole-fat dairy products is associated with the risk of developing advanced or fatal PCa in elderly men and worse survival in PCa cases. Thus, minimal intake of whole-fat dairy products may be benefi for elderly men, particularly PCa survivors. However, these results still need to be confi in other male populations.

#### Acknowledgments

Y.S. analyzed data and wrote the manuscript; J.M. supervised the analysis and edited the manuscript; and J.E.C., Y.C., W.Q., L.M., H.D.S., M.J.S., E.G., M.P., and S.L. revised the article for intellectual content. All authors read and approved the final manuscript.

#### Literature Cited

- Chan JM, Stampfer MJ, Giovannucci EL. What causes prostate cancer? A brief summary of the epidemiology. *Semin Cancer Biol.* 1998;8:263–73.
- Chan JM, Gann PH, Giovannucci EL. Role of diet in prostate cancer development and progression. *J Clin Oncol.* 2005;23:8152–60.
- Raimondi S, Mabrouk JB, Shatenstein B, Maisonneuve P, Ghadirian P. Diet and prostate cancer risk with specifi focus on dairy products and dietary calcium: a case-control study. *Prostate.* 2010;70:1054–65.
- Kurahashi N, Inoue M, Iwasaki M, Sasazuki S, Tsugane AS. Dairy product, saturated fatty acid, and calcium intake and prostate cancer in a prospective cohort of Japanese men. *Cancer Epidemiol Biomarkers Prev.* 2008;17:930–7.
- Mitrou PN, Albanes D, Weinstein SJ, Pietinen P, Taylor PR, Virtamo J, Leitzmann MF. A prospective study of dietary calcium, dairy products and prostate cancer risk (Finland). *Int J Cancer.* 2007;120:2466–73.
- Ahn J, Albanes D, Peters U, Schatzkin A, Lim U, Freedman M, Chatterjee N, Andriole GL, Leitzmann MF, Hayes RB. Dairy products, calcium intake, and risk of prostate cancer in the prostate, lung, colorectal, and ovarian cancer screening trial. *Cancer Epidemiol Biomarkers Prev.* 2007;16:2623–30.
- Kesse E, Bertrais S, Astorg P, Jaouen A, Arnault N, Galan P, Hercberg S. Dairy products, calcium and phosphorus intake, and the risk of prostate cancer: results of the French prospective SU.VI.MAX (Supplementation en Vitamines et Mineraux Antioxydants) study. *Br J Nutr.* 2006;95:539–45.
- Tseng M, Breslow RA, Graubard BI, Ziegler RG. Dairy, calcium, and vitamin D intakes and prostate cancer risk in the National Health and Nutrition Examination Epidemiologic Follow-up Study cohort. *Am J Clin Nutr.* 2005;81:1147–54.
- Chan JM, Stampfer MJ, Ma J, Gann PH, Gaziano JM, Giovannucci EL. Dairy products, calcium, and prostate cancer risk in the Physicians' Health Study. *Am J Clin Nutr.* 2001;74:549–54.
- Park SY, Murphy SP, Wilkens LR, Stram DO, Henderson BE, Kolonel LN. Calcium, vitamin D, and dairy product intake and prostate cancer risk: the Multiethnic Cohort Study. *Am J Epidemiol.* 2007;166:1259–69.
- Koh KA, Sesso HD, Paffenbarger RS Jr, Lee IM. Dairy products, calcium and prostate cancer risk. *Br J Cancer.* 2006;95:1582–5.
- Gallus S, Bravi F, Talamini R, Negri E, Montella M, Ramazzotti V, Franceschi S, Giacosa A, La Vecchia C. Milk, dairy products and cancer risk (Italy). *Cancer Causes Control.* 2006;17:429–37.
- Gao X, LaValley MP, Tucker KL. Prospective studies of dairy product and calcium intakes and prostate cancer risk: a meta-analysis. *J Natl Cancer Inst.* 2005;97:1768–77.
- Huncharek M, Muscat J, Kupelnick B. Dairy products, dietary calcium and vitamin D intake as risk factors for prostate cancer: a meta-analysis of 26,769 cases from 45 observational studies. *Nutr Cancer.* 2008;60:421–41.
- Final report on the aspirin component of the ongoing Physicians' Health Study. Steering Committee of the Physicians' Health Study Research Group. *N Engl J Med.* 1989;321:129–35.
- Hennekens CH, Buring JE, Manson JE, Stampfer M, Rosner B, Cook NR, Belanger C, LaMotte F, Gaziano JM, Ridker PM, et al. Lack of effect of long-term supplementation with beta carotene on the incidence of malignant neoplasms and cardiovascular disease. *N Engl J Med.* 1996;334:1145–9.
- USDA National Agricultural Library. USDA National Nutrient Database for Standard Reference; 2011 [cited 2012 Oct 8]. Available from: <http://ndb.nal.usda.gov>.
- Gaziano JM, Glynn RJ, Christen WG, Kurth T, Belanger C, MacFadyen J, Bubes V, Manson JE, Sesso HD, Buring JE. Vitamins E and C in the prevention of prostate and total cancer in men: the Physicians' Health Study II randomized controlled trial. *JAMA.* 2009;301:52–62.
- Chan JM, Giovannucci E, Andersson SO, Yuen J, Adami HO, Wolk A. Dairy products, calcium, phosphorus, vitamin D, and risk of prostate cancer (Sweden). *Cancer Causes Control.* 1998;9:559–66.
- Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Roy Stat Soc B Met.* 1995;57:289–300.
- Fulgioni V III, Nicholls J, Reed A, Buckley R, Kafer K, Huth P, DiRienzo D, Miller GD. Dairy consumption and related nutrient intake in African-American adults and children in the United States: continuing survey of food intakes by individuals 1994–1996, 1998, and the National Health And Nutrition Examination Survey 1999–2000. *J Am Diet Assoc.* 2007;107:256–64.
- Gonzalez CA, Riboli E. Diet and cancer prevention: contributions from the European Prospective Investigation into Cancer and Nutrition (EPIC) study. *Eur J Cancer.* 2010;46:2555–62.
- Raimondi S, Mabrouk JB, Shatenstein B, Maisonneuve P, Ghadirian P. Diet and prostate cancer risk with specific focus on dairy products and dietary calcium: a case-control study. *Prostate.* 2010;70:1054–65.
- Lanou AJ. Should dairy be recommended as part of a healthy vegetarian diet? Counterpoint. *Am J Clin Nutr.* 2009;89:S1638–42.
- Weaver CM. Should dairy be recommended as part of a healthy vegetarian diet? Point. *Am J Clin Nutr.* 2009;89:S1634–7.
- Rodriguez C, McCullough ML, Mondul AM, Jacobs EJ, Fakhrabadi-Shokoochi D, Giovannucci EL, Thun MJ, Calle EE. Calcium, dairy products, and risk of prostate cancer in a prospective cohort of United States men. *Cancer Epidemiol Biomarkers Prev.* 2003;12:597–603.
- Torniaainen S, Hedelin M, Autio V, Rasinpera H, Balter KA, Klint A, Bellocco R, Wiklund F, Stattin P, Ikonen T, et al. Lactase persistence, dietary intake of milk, and the risk for prostate cancer in Sweden and Finland. *Cancer Epidemiol Biomarkers Prev.* 2007;16:956–61.
- Park Y, Mitrou PN, Kipnis V, Hollenbeck A, Schatzkin A, Leitzmann MF. Calcium, dairy foods, and risk of incident and fatal prostate cancer: the NIH-AARP Diet and Health Study. *Am J Epidemiol.* 2007;166:1270–9.
- Gann PH. Risk factors for prostate cancer. *Rev Urol.* 2002;4 Suppl 5: S3–10.
- U.S. Preventive Services Task Force. Screening for prostate cancer; 2012 [cited 2012 Oct 8]. Available from: <http://www.uspreventiveservices-taskforce.org/prostatecancerscreening.htm>.
- Chan JM, Holick CN, Leitzmann MF, Rimm EB, Willett WC, Stampfer MJ, Giovannucci EL. Diet after diagnosis and the risk of prostate cancer progression, recurrence, and death (United States). *Cancer Causes Control.* 2006;17:199–208.
- Pettersson A, Kasperzyk JL, Kenfield SA, Richman EL, Chan JM, Willett WC, Stampfer MJ, Mucci LA, Giovannucci EL. Milk and dairy consumption among men with prostate cancer and risk of metastases

and prostate cancer death. *Cancer Epidemiol, Biomarkers Prev.* 2012;21:428–36.

33. Torfadottir JE, Steingrimsdottir L, Mucci L, Aspelund T, Kasperzyk JL, Olafsson O, Fall K, Tryggvadottir L, Harris TB, Launer L, et al. Milk intake in early life and risk of advanced prostate cancer. *Am J Epidemiol.* 2012;175:144–53.
34. Giovannucci E. Dietary influences of 1,25(OH)<sub>2</sub> vitamin D in relation to prostate cancer: a hypothesis. *Cancer Causes Control.* 1998;9:567–82.
35. Giovannucci E, Liu Y, Stampfer MJ, Willett WC. A prospective study of calcium intake and incident and fatal prostate cancer. *Cancer Epidemiol Biomarkers Prev.* 2006;15:203–10.
36. Allen NE, Key TJ, Appleby PN, Travis RC, Roddam AW, Tjonneland A, Johnsen NF, Overvad K, Linseisen J, Rohrmann S, et al. Animal foods, protein, calcium and prostate cancer risk: the European Prospective Investigation into Cancer and Nutrition. *Br J Cancer.* 2008;98:1574–81.
37. Hellgren LI. Phytanic acid: an overlooked bioactive fatty acid in dairy fat? *Ann N Y Acad Sci.* 2010;1190:42–9.
38. Price AJ, Allen NE, Appleby PN, Crowe FL, Jenab M, Rinaldi S, Slimani N, Kaaks R, Rohrmann S, Boeing H, et al. Plasma phytanic acid concentration and risk of prostate cancer: results from the European Prospective Investigation into Cancer and Nutrition. *Am J Clin Nutr.* 2010;91:1769–76.
39. Luo J, Zha S, Gage WR, Dunn TA, Hicks JL, Bennett CJ, Ewing CM, Platz EA, Ferdinandusse S, Wanders RJ, et al. Alpha-methylacyl-CoA racemase: a new molecular marker for prostate cancer. *Cancer Res.* 2002;62:2220–6.
40. Newmark HL, Heaney RP. Dairy products and prostate cancer risk. *Nutr Cancer.* 2010;62:297–9.
41. McGreevy KM, Hoel BD, Lipsitz SR, Hoel DG. Impact of nutrients on insulin-like growth factor-I, insulin-like growth factor binding protein-3 and their ratio in African American and white males. *Public Health Nutr.* 2007;10:97–105.
42. Ma J, Giovannucci E, Pollak M, Chan JM, Gaziano JM, Willett W, Stampfer MJ. Milk intake, circulating levels of insulin-like growth factor-I, and risk of colorectal cancer in men. *J Natl Cancer Inst.* 2001;93:1330–6.
43. Qin LQ, Wang PY, Kaneko T, Hoshi K, Sato A. Estrogen: one of the risk factors in milk for prostate cancer. *Med Hypotheses.* 2004;62:133–42.
44. Giovannucci E, Rimm EB, Liu Y, Willett WC. Height, predictors of C-peptide and cancer risk in men. *Int J Epidemiol.* 2004;33:217–25.

## Original Contribution

# Association of Type 2 Diabetes Susceptibility Variants With Advanced Prostate Cancer Risk in the Breast and Prostate Cancer Cohort Consortium

Mitchell J. Machiela, Sara Lindström, Naomi E. Allen, Christopher A. Haiman, Demetrius Albanes, Aurelio Barricarte, Sonja I. Berndt, H. Bas Bueno-de-Mesquita, Stephen Chanock, J. Michael Gaziano, Susan M. Gapstur, Edward Giovannucci, Brian E. Henderson, Eric J. Jacobs, Laurence N. Kolonel, Vittorio Krogh, Jing Ma, Meir J. Stampfer, Victoria L. Stevens, Daniel O. Stram, Anne Tjønneland, Ruth Travis, Walter C. Willett, David J. Hunter, Loic Le Marchand, and Peter Kraft\*

\* Correspondence to Dr. Peter Kraft, Department of Epidemiology, Harvard School of Public Health, 655 Huntington Avenue, Building II, Room 207, Boston, MA 02115 (e-mail: pkraft@hsph.harvard.edu).

Initially submitted September 22, 2011; accepted for publication March 26, 2012.

Observational studies have found an inverse association between type 2 diabetes (T2D) and prostate cancer (PCa), and genome-wide association studies have found common variants near 3 loci associated with both diseases. The authors examined whether a genetic background that favors T2D is associated with risk of advanced PCa. Data from the National Cancer Institute's Breast and Prostate Cancer Cohort Consortium, a genome-wide association study of 2,782 advanced PCa cases and 4,458 controls, were used to evaluate whether individual single nucleotide polymorphisms or aggregations of these 36 T2D susceptibility loci are associated with PCa. Ten T2D markers near 9 loci (*NOTCH2*, *ADCY5*, *JAZF1*, *CDKN2A/B*, *TCF7L2*, *KCNQ1*, *MTNR1B*, *FTO*, and *HNFB*) were nominally associated with PCa ( $P < 0.05$ ); the association for single nucleotide polymorphism rs757210 at the *HNFB* locus was significant when multiple comparisons were accounted for (adjusted  $P = 0.001$ ). Genetic risk scores weighted by the T2D log odds ratio and multilocus kernel tests also indicated a significant relation between T2D variants and PCa risk. A mediation analysis of 9,065 PCa cases and 9,526 controls failed to produce evidence that diabetes mediates the association of the *HNFB* locus with PCa risk. These data suggest a shared genetic component between T2D and PCa and add to the evidence for an interrelation between these diseases.

carcinoma; diabetes mellitus, type 2; genetic predisposition to disease; genetics; genome-wide association study; humans; polymorphism, single nucleotide; prostatic neoplasms

Abbreviations: BPC3, Breast and Prostate Cancer Cohort Consortium; CI, confidence interval; GRS, genetic risk score; OR, odds ratio; PCa, prostate cancer; SNP, single nucleotide polymorphism; T2D, type 2 diabetes.

Prostate cancer (PCa) and type 2 diabetes (T2D) are two of the most common chronic diseases afflicting the US aging male population (1, 2). Observational studies have consistently shown an apparent inverse association between T2D and risk of PCa, with meta-analysis risk ratios ranging from 0.84 to 0.91 (3, 4). The reduction in PCa risk has been reported to increase with years since T2D diagnosis, with men who have had T2D for more than 15 years being at a 22% reduced hazard of PCa (5). The association is

poorly understood, with one hypothesis suggesting that the metabolic status of men with T2D could move gradually from hyperinsulinemia to endogenous insulin deficiency, which could mitigate the oncogenic action of insulin in the prostate (6, 7).

Recently, 3 shared genomic regions for T2D and PCa have been highlighted. The first region, located on chromosome 17, is in intron 2 of *HNFB*, formerly known as *TCF2*. The major allele A of rs4430796 is positively

associated with PCa risk (odds ratio (OR) = 1.22) and inversely associated with risk of T2D (OR = 0.91) (8–10). The second region is located on chromosome 7 near the *JAZF1* locus, where the major allele G of rs10486567 is inversely associated with risk of PCa (aggressive PCa: OR = 0.89; nonaggressive PCa: OR = 0.74) (11), whereas the minor allele G of rs864745 is positively associated with T2D (OR = 1.10) (12). *THADA* is the third region, located on chromosome 2, with the minor allele A of rs1465618 being associated with PCa (OR = 1.08) (13) and the major allele T of rs7578597 associated with T2D (OR = 1.15) (12). However, the single nucleotide polymorphisms (SNPs) for T2D and PCa in the *JAZF1* and *THADA* regions are weakly linked, with  $R^2$  values of 0.03 and 0.02, respectively. It is not clear that these associations are driven by the same haplotype (14, 15).

Stevens et al. (16) investigated the T2D-PCa relation further and concluded that diabetic status did not mediate the observed relation between the *HNF1B* and *JAZF1* gene variants and PCa risk. In the Atherosclerosis Risk in Communities cohort, Meyer et al. (17) examined the relation of T2D-associated variants with risk of PCa and found that 4 of 13 T2D SNPs were nominally associated with PCa, which provides additional evidence that some of the T2D-PCa association could be driven by shared genetic factors. Another study by Pierce et al. (18) evaluated the ability of risk scores, consisting of 18 replicated T2D risk variants, to predict PCa risk and concluded that persons with increased genetic susceptibility to T2D have a reduced risk of PCa. However, in a recent study of 5 racial/ethnic groups in the Multiethnic Cohort and PAGE (Population Architecture using Genomics and Epidemiology), Waters et al. (19) found no association between T2D risk variants, either individually or in risk scores, and PCa risk.

With a large sample size and an expanded set of recently published T2D susceptibility loci, we aimed to investigate whether and to what extent individual T2D risk variants and aggregations of T2D replicated risk variants are associated with PCa risk. We used novel approaches to test both whether these risk variants are inversely associated with PCa risk in accordance with the inverse relation observed between T2D and PCa in observational studies and, more generally, whether these T2D loci are associated with PCa risk without regard to directionality of association. Additionally, using causal inference methods, our study attempted to more definitively investigate the potential for mediation of the effect of *HNF1B* on PCa risk through T2D phenotype.

## MATERIALS AND METHODS

Genotyping data for PCa cases and controls came from the National Cancer Institute's Breast and Prostate Cancer Cohort Consortium (BPC3). The BPC3 is a consortium of prospective cohort studies, with contributors including the Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study (20), the American Cancer Society Cancer Prevention Study II Nutrition Cohort (21), the European Prospective Investigation into Cancer and Nutrition (22), the Health Professionals Follow-up Study, the Melbourne Collaborative Cohort Study (23), the Multiethnic Cohort Study (24),

the Physicians' Health Study, and the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial (25). In total, 9,065 PCa cases and 9,526 controls comprised the PCa nested case-control study. Diabetes phenotype was self-reported at study baseline, with data available for 96.7% of BPC3 participants. A genome-wide association scan was conducted on a subset of 2,782 European cases with advanced disease and 4,458 controls with European ancestry. Advanced PCa was defined as PCa cases that had either a high histologic grade (Gleason score  $\geq 8$ ) or extraprostatic extension (stage C/D). All controls were free of PCa at the time of selection and were sampled from the same cohort as the cases. Controls were age-matched to cases, and study indicator variables were used to adjust for sampling differences between studies. Informed consent was received from all study participants, and all study protocols were reviewed by the institutional review boards of the National Cancer Institute and each participating study center.

A literature search was conducted to find robustly replicated disease susceptibility loci that are associated with T2D at genome-wide significance levels ( $P < 5 \times 10^{-8}$ ). In total, 36 independent autosomal loci associated with T2D were identified, and published T2D risk alleles and odds ratios were extracted (9, 10, 12, 26–36).

Individual association tests were carried out for each T2D SNP with PCa risk in the BPC3 genome-wide association study (37). Quality control filters were used to remove samples with heterozygosity, underperforming samples or markers, markers with genotype frequencies that significantly departed from Hardy-Weinberg equilibrium, and subjects with significant evidence of non-European ancestry or sample structure. Of the 36 T2D SNPs, 19 were not directly genotyped on the Illumina HumanHap610 Quad Arrays (Illumina, San Diego, California) and were therefore imputed with MACH (<http://www.sph.umich.edu/csg/abecasis/MaCH/>) (38). MACH references the HapMap (<http://hapmap.ncbi.nlm.nih.gov/>) CEU population (Utah residents with Northern and Western European ancestry from the Centre d'Etude du Polymorphisme Humain (CEPH) collection) to infer expected genotype counts for each marker locus. MACH quality scores and  $R^2$  values were more than 0.85 and 0.75, respectively, for all imputed SNPs. Logistic regression models were used to test for T2D SNP associations with PCa risk. The number of T2D risk alleles was used as the exposure, and adjustment was made for cohort (indicator variables). A nominal association  $P$  value of 0.05 was used to assess whether T2D markers exhibited more significant associations with PCa than would be expected by chance. Additional binomial and permutation tests (39) (10,000 permutations) were carried out to test for a relation in risk allele directionality and significant departures of the PCa association statistics from the null distribution, respectively.

The T2D SNPs were combined to form a genetic risk score (GRS) using the `--score` command in PLINK (40). The GRS was calculated in two ways. The first method, referred to here as the count method, involved summing the number of T2D risk alleles at each locus (0, 1, or 2) and then summing across all T2D loci. This count method is an additive model that weights each locus equally and assumes no gene-gene interactions. The second method, referred to

here as the weighted method, uses the log odds ratio of the published T2D loci to weight the sum of T2D risk alleles at each locus and then sums across all T2D loci. The weighted method is an additive model that weights each locus in accordance with the T2D literature and assumes no gene-gene interactions. The rationale for weighting is to create a score that is the best GRS for T2D and therefore can be used as an instrument for testing an association with PCa. For each GRS method, we included the GRS as a predictor in a logistic regression model with PCa case-control status as the outcome, and we adjusted for cohort with an indicator variable. Cohort-specific associations were also calculated.

Additionally, multilocus linear kernel tests were used to assess the joint relation between the 36 T2D variants and PCa risk. These linear models allow associations of multiple genetic loci to be tested simultaneously with one test statistic (41) and have been generalized for dichotomous outcomes (42). Unlike the GRS methods, these tests require no prespecification of risk allele directionality (i.e., that the risk allele is associated with increased risk of T2D and decreased risk of PCa).

The *HNF1B* locus was the only T2D locus significantly associated with PCa risk after adjustment for multiple comparisons, so it was carried forward for mediation analysis to evaluate whether T2D phenotype is a potential mediator of the relation between *HNF1B* and PCa. We used an expanded set of data on 9,065 PCa cases (including nonaggressive cases) and 9,526 controls from the BPC3 (43) with self-reported information on diabetes phenotype. Data on rs7501939 at *HNF1B* were generated as part of a previous project characterizing known PCa loci; this SNP is in high linkage disequilibrium with rs757210 ( $R^2 = 0.81$ ). This was the only T2D risk marker typed in the larger BPC3 data set. To assess mediation, we used the mediation framework proposed by Baron and Kenny (44), extended into the counterfactual framework by VanderWeele and Vansteelandt (45) as direct and indirect effects, and further generalized for use with dichotomous intermediate and outcome. This framework for mediation analysis is flexible to an interaction between exposure and an intermediate factor, has a causal interpretation, and can assess mediation on both the multiplicative and additive scales. Assessing mediation in this manner involved fitting both an outcome model and a mediator model. The outcome model was a logistic regression model that modeled PCa as the outcome, included parameters for the T2D variant of interest and diabetes phenotype, and adjusted for potential confounders of the exposure-outcome and intermediate-outcome relations, including cohort indicator, age at baseline, and body mass index (weight (kg)/height (m)<sup>2</sup>). The mediator model was a logistic regression model that modeled diabetes phenotype as the outcome, included a parameter for the T2D variant of interest, and controlled for potential confounders, including cohort indicator, age at baseline, and body mass index. In the mediator model, the case-control nature of the BPC3 needed to be accounted for to obtain consistent effect estimates. This was accomplished by fitting the model only in the PCa controls, who represent the study's base population, and assuming a rare outcome. Once both the outcome and mediator models were fitted, parameter estimates were

used to calculate direct and indirect (mediated) effects by which to assess mediation (45).

The PCa study was conducted between May and August of 2011. All statistical analyses were carried out in SAS 9.1 (SAS Institute Inc., Cary, North Carolina) and R 2.11.1 (R Foundation for Statistical Computing, Vienna, Austria).

## RESULTS

Results from the individual association tests showed that 10 of the 36 T2D markers had a *P* value less than 0.05 for association with PCa, significantly more than the 1.8 markers that would be expected by chance ( $P = 7.5 \times 10^{-6}$ ) (Table 1). These markers include the *HNF1B* and *JAZF1* loci, as well as *NOTCH2*, *ADCY5*, *CDKN2A/B*, *TCF7L2*, *MTNR1B*, *FTO*, and 2 independent loci at *KCNQ1* (Table 1). After permutation adjustment for multiple comparisons, only *HNF1B* remained significant (adjusted  $P = 0.001$ ). Small fluctuations in effect estimates of  $\leq 3\%$  were observed when adjustment for diabetes status was made in the models, with overall conclusions remaining the same (results not shown). We observed an inflation in the observed *P* values for these 36 SNPs ( $\lambda_{GC} = 2.0$ ; Figure 1). When the observed  $\lambda_{GC}$  was compared with the distribution of permutation  $\lambda_{GC}$  values, the observed  $\lambda_{GC}$  was significantly elevated ( $P = 0.03$ ), which indicated that the distribution of association *P* values was significantly lower than expected.

We used exact binomial tests to assess whether significantly more T2D risk alleles were inversely associated with PCa risk than would be expected by chance. By chance alone, 1.8 of the 36 markers would be expected to be significant, of which, under the null, 0.9 would be expected to be significantly associated with increased risk of PCa and 0.9 would be expected to be significantly associated with decreased risk of PCa. In our data, we observed 2 T2D loci that were significantly associated with increased PCa risk, which did not differ statistically from the 0.9 loci expected by chance ( $P = 0.23$ ). However, the 8 T2D loci we observed to be significantly associated with reduced risk of PCa were significantly more than the 0.9 that would be expected by chance ( $P = 2.45 \times 10^{-6}$ ), which indicates that more T2D risk alleles than expected are associated with reduced risk of PCa.

Associations for GRS using both the unweighted count and the weighted log odds method are shown in Table 2. The risk score for the unweighted count did not show evidence for an association of these genetic variants with PCa risk. However, a significant association was observed for the weighted log odds method when *HNF1B* was both included in ( $P = 0.002$ ) and excluded from ( $P = 0.015$ ) the GRS. No changes in results were observed when we adjusted for diabetes status in the models (results not shown). Study-specific analyses showed that the log odds-weighted GRS was statistically significant only in the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial, although the test for heterogeneity indicated no significant departures from homogeneity ( $P = 0.60$ ).

The multilocus kernel test that jointly tested for a PCa association with all 36 T2D loci without specifying weight

Table 1. Individual Associations of 36 Independent Type 2 Diabetes Susceptibility Variants With Prostate Cancer Risk in the Breast and Prostate Cancer Cohort Consortium<sup>a</sup>

Chromosome	Reported Gene(s)	Single Nucleotide Polymorphism	Genotyped? <sup>b</sup>	Type 2 Diabetes Risk Allele	Frequency of Risk Allele	Odds Ratio <sup>c</sup>	95% Confidence Interval	P Value	Adjusted P Value
1	<i>NOTCH2</i>	rs10923931	No	T	0.11	0.86 <sup>d</sup>	0.76, 0.96	0.008*	0.255
1	<i>PROX1</i>	rs340874	Yes	C	0.52	1.01	1.08, 0.94	0.845	1.000
2	<i>GCKR</i>	rs780094	Yes	C	0.61	0.98 <sup>d</sup>	1.05, 0.91	0.498	1.000
2	<i>THADA</i>	rs7578597	Yes	T	0.91	1.03	1.16, 0.91	0.644	1.000
2	<i>BCL11A</i>	rs243021	Yes	A	0.47	1.02	0.95, 1.10	0.511	1.000
2	<i>IRS1</i>	rs2943641	Yes	C	0.64	0.95 <sup>d</sup>	1.02, 0.88	0.140	0.995
3	<i>PPARG</i>	rs1801282	No	C	0.86	0.96 <sup>d</sup>	1.07, 0.87	0.465	1.000
3	<i>ADAMTS9</i>	rs4607103	No	C	0.76	0.99 <sup>d</sup>	1.08, 0.91	0.853	1.000
3	<i>ADCY5</i>	rs11708067	No	A	0.78	0.91 <sup>d</sup>	0.99, 0.84	0.028*	0.630
3	<i>IGF2BP2</i>	rs4402960	Yes	T	0.32	1.03	0.95, 1.11	0.456	1.000
4	<i>WFS1</i>	rs10010131	No	G	0.60	1.00	1.07, 0.93	0.924	1.000
5	<i>ZBED3</i>	rs4457053	No	G	0.29	1.02	0.94, 1.10	0.672	1.000
6	<i>CDKAL1</i>	rs7754840	Yes	C	0.32	1.04	0.97, 1.13	0.270	1.000
7	<i>DGKB</i>	rs2191349	No	T	0.52	1.00	1.07, 0.93	0.945	1.000
7	<i>JAZF1</i>	rs864745	No	T	0.50	1.08	1.16, 1.01	0.033*	0.694
7	<i>GCK</i>	rs4607517	Yes	A	0.15	1.06	0.96, 1.16	0.256	1.000
7	<i>KLF14</i>	rs972283	No	G	0.53	1.02	1.09, 0.95	0.627	1.000
8	<i>TP53INP1</i>	rs896854	Yes	T	0.51	1.02	1.09, 0.95	0.668	1.000
8	<i>SLC30A8</i>	rs13266634	Yes	C	0.68	1.00	1.08, 0.93	0.963	1.000
9	<i>CDKN2A/B</i>	rs10811661	No	T	0.82	0.91 <sup>d</sup>	1.00, 0.83	0.045*	0.809
9	<i>TLE4</i>	rs13292136	No	C	0.93	0.93 <sup>d</sup>	1.07, 0.81	0.312	1.000
10	<i>CDC123/CAMK1D</i>	rs12779790	No	G	0.18	1.06	0.97, 1.16	0.206	1.000
10	<i>HHEX/IDE</i>	rs1111875	Yes	C	0.58	1.01	1.09, 0.94	0.713	1.000
10	<i>TCF7L2</i>	rs7903146	Yes	T	0.28	0.90 <sup>d</sup>	0.83, 0.97	0.009*	0.276
11	<i>KCNQ1</i>	rs231362	No	G	0.50	0.92 <sup>d</sup>	0.86, 0.98	0.014*	0.393
11	<i>KCNQ1</i>	rs2237892	Yes	C	0.94	0.85 <sup>d</sup>	0.98, 0.74	0.030*	0.659
11	<i>KCNJ11</i>	rs5215	Yes	T	0.61	0.99 <sup>d</sup>	1.06, 0.92	0.719	1.000
11	<i>CENTD2</i>	rs1552224	Yes	A	0.83	1.00	1.10, 0.91	0.963	1.000
11	<i>MTNR1B</i>	rs10830963	No	G	0.28	1.10	1.01, 1.19	0.023*	0.561
12	<i>HMGA2</i>	rs1531343	No	C	0.10	0.98 <sup>d</sup>	0.88, 1.10	0.764	1.000
12	<i>TSPAN8/LGR5</i>	rs7961581	No	C	0.26	1.05	0.97, 1.13	0.259	1.000
12	<i>HNF1A/TCF1</i>	rs7957197	No	T	0.80	0.96 <sup>d</sup>	1.05, 0.88	0.346	1.000
15	<i>ZFAND6</i>	rs11634397	No	G	0.66	1.04	1.12, 0.96	0.346	1.000
15	<i>PRC1</i>	rs8042680	Yes	A	0.32	1.04	0.97, 1.12	0.286	1.000
16	<i>FTO</i>	rs9939609	No	A	0.40	0.93 <sup>d</sup>	0.86, 1.00	0.041*	0.775
17	<i>HNF1B/TCF2</i>	rs757210	Yes	T	0.35	0.85 <sup>d</sup>	0.79, 0.92	3e <sup>-05</sup> *	0.001 <sup>e</sup>

Abbreviations: CI, confidence interval; OR, odds ratio; T2D, type 2 diabetes; RA, risk allele; SNP, single nucleotide polymorphism.

\*  $P < 0.05$ .

<sup>a</sup> Association tests were carried out in the Breast and Prostate Cancer Cohort Consortium using a log-additive genetic model with adjustment made for cohort indicators.

<sup>b</sup> Indicates whether or not variants were genotyped. Variants that were not directly genotyped were imputed.

<sup>c</sup> Odds ratio for the increase in prostate cancer risk associated with a 1-unit increase in the number of type 2 diabetes risk alleles carried at each locus.

<sup>d</sup> Association for prostate cancer was in the inverse direction.

<sup>e</sup> Significant after permutation correction for multiple testing.

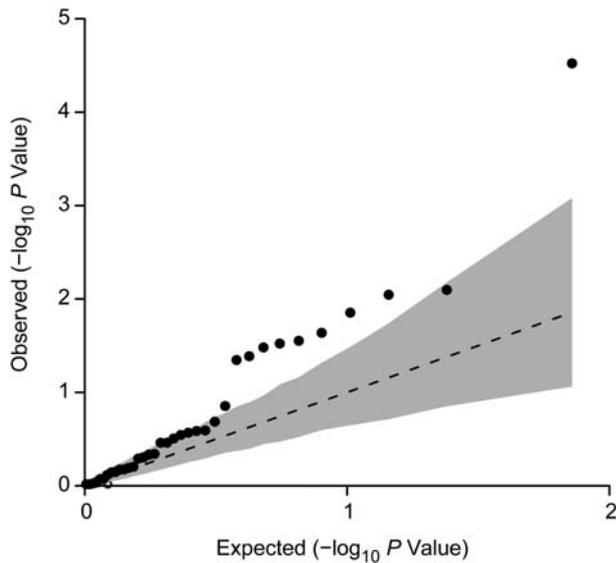


Figure 1. Quantile-quantile plot comparing the uniformly distributed  $-\log_{10} P$  values for the 36 type 2 diabetes (T2D) susceptibility markers with  $-\log_{10} P$  values observed in the Breast and Prostate Cancer Cohort Consortium data set when the authors tested for an association with prostate cancer (PCa) risk by means of a Wald test. The dotted line shows the expected  $-\log_{10} P$  value distribution. The black points represent observed  $P$  values for the association of each T2D locus with PCa risk. The gray region is the 95% confidence interval for 10,000 permutations. The inflation index ( $\lambda_{GC}$ ) of 1.95 is significantly elevated ( $P = 0.02$ ), which indicates an overall inflation in association  $P$  values but gives no information about the directionality of association between the T2D variants and PCa risk.

or directionality of risk alleles was statistically significant ( $P = 0.0001$ ). When *HNF1B* was removed from the list of included markers and the remaining 35 markers were fitted, the  $P$  value was attenuated but remained significant ( $P = 0.01$ ), which indicated that a substantial portion of the association was a result of the *HNF1B* locus but that other T2D loci were associated with PCa as well.

We conducted mediation analyses for the *HNF1B* locus to investigate whether the locus had effects that act directly on PCa risk or whether the effects of the locus were mediated through diabetes phenotype (Table 3). The outcome model produced significant evidence for an association between *HNF1B* and PCa risk (OR = 0.83, 95% confidence interval (CI): 0.79, 0.86;  $P = 6.37 \times 10^{-19}$ ) and an association between diabetes phenotype and PCa risk (OR = 0.76, 95% CI: 0.66, 0.87;  $P = 8.13 \times 10^{-5}$ ). The mediator model indicated that the minor T allele of rs7501939 was not statistically significantly associated with an increased risk of diabetes among the 9,526 PCa controls (OR = 1.10, 95% CI: 0.97, 1.25;  $P = 0.14$ ), although the per-allele odds ratio for association with T2D was consistent with previous reports (8–10). When these results were combined together, the estimated direct effect of *HNF1B* on PCa risk was statistically significant (OR = 0.83, 95% CI: 0.79, 0.86;  $P = 1.02 \times 10^{-18}$ ), but the mediated (indirect) effect through diabetes phenotype was

nonsignificant (OR = 1.00, 95% CI: 1.00, 1.00;  $P = 0.71$ ). These results are in agreement with the standard mediation analysis, which produced an insignificant 0.5% change in the parameter estimate for the effect of *HNF1B* when diabetes status was included as a covariate.

## DISCUSSION

Our study suggests that genetic variants associated with T2D are also associated with PCa risk. Ten of 36 T2D susceptibility markers were nominally associated with PCa risk at *NOTCH2*, *ADCY5*, *JAZF1*, *CDKN2A/B*, *TCF7L2*, *KCNQ1*, *MTNR1B*, *FTO*, and *HNF1B*, although only the *HNF1B* locus remained significantly associated with PCa risk after adjustment for multiple testing. However, log odds ratio-weighted GRS and kernel machine models also were associated with PCa risk both with and without inclusion of the *HNF1B* locus, which suggests that other genetic variants associated with T2D risk also contribute to PCa risk. Finally, mediation analysis provided insufficient evidence that the association of the *HNF1B* locus with PCa risk is mediated through diabetes phenotype.

Our study adds to the evidence that a genetic background favorable to the development of T2D is associated with PCa risk. The *HNF1B* locus was most strongly associated with PCa risk in this analysis and accounted for some but not all of the association between the T2D variants and PCa risk in the GRS and the kernel regression. The noted inflation in our association  $P$  values for other T2D SNPs is consistent with what others have observed (17, 18) and indicates that more germline variants are held in common between T2D and PCa than would be expected by chance.

Our study's large sample size and recently published T2D susceptibility loci permitted us to detect potentially novel genetic relations between T2D and PCa that have not been reported previously. Seven loci (*NOTCH2*, *ADCY5*, *CDKN2A/B*, *TCF7L2*, *KCNQ1*, *MTNR1B*, and *FTO*) not previously associated with PCa at genome-wide significance levels were seen as nominally associated in our study, one of which (*FTO*) was also reported by Pierce et al. (18). Four of these loci (*CDKN2A/B*, *TCF7L2*, *KCNQ1*, and *MTNR1B*) are associated with altered beta cell dysfunction or impaired insulin release and could result in less insulin production, thus blunting insulin effects in increasing PCa risk (46). Additionally, our second most highly associated locus, the *NOTCH2* locus ( $P = 0.008$ ; permutation  $P = 0.26$ ), is of interest. *NOTCH2* is a member of the *NOTCH* family of receptors, which modulate cellular differentiation, proliferation, and apoptosis (47). The locus has been reported to be associated with both T2D and breast cancer (48, 49). Evidence from gene expression data indicates that *NOTCH2* is expressed in developing prostate stroma and that *NOTCH* signaling affects stromal survival only in the presence of testosterone (50). Therefore, the regulatory ability of *NOTCH2* and its sensitivity to the presence of testosterone might be important in prostate carcinogenesis, although additional studies are needed to investigate this further.

Our use of GRS and kernel machine models allowed us to investigate the cumulative effect of T2D susceptibility

Table 2. Individual Cohort and Combined Results for Unweighted and Log Odds Ratio-Weighted Type 2 Diabetes Genetic Risk Score in the Breast and Prostate Cancer Cohort Consortium<sup>a</sup>

Cohort	No. in Cohort	No. of Cases	No. of Total <sup>b</sup>	Mean <sup>c</sup>		GRS			GRS (- <i>HNF1B</i> ) <sup>d</sup>			
				Cases	Controls	OR	95% CI	P Value	OR	95% CI	P Value	
Unweighted count												
ATBC	1,490	245	72	36.48	36.44	1.00	0.97, 1.04	0.894	1.00	0.97, 1.04	0.841	
CPSII	1,258	636	72	37.48	37.55	1.00	0.97, 1.03	0.740	1.00	0.97, 1.03	0.839	
EPIC	857	431	72	37.47	37.66	0.99	0.95, 1.02	0.460	1.00	0.96, 1.04	0.984	
HPFS	418	214	72	37.70	37.47	1.02	0.97, 1.07	0.539	1.02	0.97, 1.08	0.419	
MEC	503	244	72	37.80	37.89	0.99	0.95, 1.04	0.779	1.00	0.96, 1.05	0.936	
PHS	553	298	72	37.59	37.81	0.99	0.95, 1.03	0.521	1.00	0.95, 1.04	0.800	
PLCO	2,161	714	72	37.36	37.64	0.98	0.96, 1.00	0.111	0.98	0.96, 1.01	0.191	
Combined <sup>e</sup>	7,240	2,782	72	37.42	37.31	0.99	0.98, 1.00	0.168	1.00	0.98, 1.01	0.534	
Weighted log OR												
ATBC	1,490	245	8.16	4.33	4.34	0.93	0.68, 1.29	0.675	0.94	0.68, 1.30	0.718	
CPSII	1,258	636	8.16	4.45	4.47	0.89	0.69, 1.14	0.358	0.90	0.70, 1.16	0.416	
EPIC	857	431	8.16	4.45	4.47	0.90	0.67, 1.20	0.460	1.01	0.75, 1.36	0.961	
HPFS	418	214	8.16	4.49	4.46	1.11	0.73, 1.68	0.635	1.17	0.76, 1.80	0.481	
MEC	503	244	8.16	4.49	4.54	0.78	0.53, 1.15	0.215	0.83	0.56, 1.23	0.352	
PHS	553	298	8.16	4.45	4.52	0.76	0.53, 1.07	0.118	0.80	0.56, 1.15	0.232	
PLCO	2,161	714	8.16	4.43	4.49	0.74	0.61, 0.91	0.004	0.76	0.62, 0.93	0.008	
Combined <sup>e</sup>	7,240	2,782	8.16	4.44	4.45	0.84	0.75, 0.94	0.002	0.87	0.78, 0.97	0.015	

Abbreviations: ATBC, Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study; CI, confidence interval; CPSII, American Cancer Society Cancer Prevention Study II Nutrition Cohort; EPIC, European Prospective Investigation into Cancer and Nutrition; GRS, genetic risk score; HPFS, Health Professionals Follow-up Study; MEC, Multiethnic Cohort Study; OR, odds ratio; PCa, prostate cancer; PHS, Physicians' Health Study; PLCO, Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial; T2D, type 2 diabetes.

<sup>a</sup> Logistic regression models were used to regress GRS on risk of PCa.

<sup>b</sup> Total indicates the maximum bound for the respective GRS, with a value close to this total indicating high genetic predisposition for T2D.

<sup>c</sup> Mean GRS was calculated for PCa cases and PCa controls.

<sup>d</sup> *HNF1B* was excluded from the GRS and included as a separate covariate.

<sup>e</sup> For combined estimates, cohort indicators were added to adjust for cohort effects.

variants on PCa risk. Although another study was successful in showing an association between unweighted T2D

Table 3. Mediation Analysis for the Association Between *HNF1B* (rs7501939) and Prostate Cancer With Diabetes Phenotype as a Potential Intermediate in the Breast and Prostate Cancer Cohort Consortium<sup>a</sup>

	Odds Ratio	95% Confidence Interval	P Value
<i>HNF1B</i> -T2D association	1.10	0.97, 1.25	0.14
T2D-prostate cancer association	0.76	0.66, 0.87	$8.13 \times 10^{-05}$
Natural indirect effect	1.00	1.00, 1.00	0.71
Natural direct effect	0.83	0.79, 0.86	$1.02 \times 10^{-18}$
Total effect	0.83	0.79, 0.86	$6.37 \times 10^{-19}$

Abbreviation: T2D, type 2 diabetes.

<sup>a</sup> All analyses were conducted in the Breast and Prostate Cancer Cohort Consortium and were adjusted for cohort indicator, age at baseline (years), and body mass index (weight (kg)/height (m)<sup>2</sup>).

GRS and PCa (18), our study did not find a relation between unweighted T2D risk scores and PCa. A potential explanation for our lack of association is that with the most recent T2D loci added to our risk score, including T2D variants found through meta-analyses with lower-than-average effect sizes, the number of SNPs doubled, and the range of effect estimates for each variant might have widened. Our study did find a significant association between the log odds-weighted T2D risk scores and PCa. This association was significant when the *HNF1B* locus was both included in and excluded from the GRS. Although one of the larger cohorts, the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial, seems to have been responsible for most of this association, a test of heterogeneity indicated that there was no significant evidence for heterogeneity. The fact that the log odds ratio-weighted GRS was significant and the unweighted risk score was insignificant indicates that some T2D variants could have a stronger influence on PCa risk than others. The GRS approach makes the assumption that all T2D loci included in the GRS have T2D risk alleles that function in the same direction when PCa risk is considered. This might not be the case, with some T2D-associated loci

possibly having the same rather than the (expected) opposite direction of effect on PCa. Multilocus kernel tests allowed us to assess the cumulative effect of these 36 T2D variants on PCa risk without requiring an assumption about risk allele directionality. Results from the multilocus kernel tests indicated that the 36 T2D variants were significantly associated with PCa risk when *HNF1B* was both included in and excluded from the models, which suggests that common pathways could be involved in both T2D and PCa.

A potential limitation of this study is that information on diabetes phenotype was self-reported (43). However, previous studies have shown that self-reporting of diabetes has up to 97% agreement with medical records (51, 52). Another limitation is that we could not differentiate between cases of type 1 diabetes and T2D, although the median age (62 years; interquartile range, 55–70) and ethnicity of our study population were such that the majority of diabetes cases were likely to be T2D (53). Furthermore, BPC3 data on T2D status were available only at baseline, and although this could have resulted in underestimation of the true prevalence of diabetes in our study population, it did guard against potential reverse causality.

Our study showed a highly significant inverse relation between T2D and PCa. The estimate was adjusted for body mass index, age at baseline, and cohort indicator and is unlikely to be due to chance or uncontrolled bias. To our knowledge, this is the largest case-control study in which this inverse association has been examined, and our estimate (OR = 0.76) is comparable to, albeit slightly stronger than, the point estimates reported in meta-analyses and other studies, including prior reports from 2 cohorts in the BPC3 (i.e., relative risks ranged from 0.84 to 0.91) (3–5, 54).

We further assessed the potential for T2D phenotype to mediate the effect of *HNF1B* with PCa risk. Results indicated a highly significant direct association between *HNF1B* and PCa risk, but there was no significant evidence for an indirect association. Although other investigators have observed a significant relation between *HNF1B* and T2D risk (8, 9), we did not, which indicates that our sample set might have lacked sufficient statistical power to detect this effect. The lack of a mediation role for diabetes phenotype in the *HNF1B*-PCa association has been reported elsewhere in a smaller subset of the BPC3 data (16), although larger studies are needed to more definitively rule out the potential for mediation.

The majority of our analysis, excluding the mediation analysis, was conducted on data from a genome-wide association study of advanced PCa. Although there is concern that results from our study might not be generalizable to other subtypes of PCa, the overwhelming number of similarities between our analysis and others indicates that T2D risk variants have a similar effect on advanced PCa risk and on total PCa risk. This is in agreement with association studies comparing PCa germline variants that show very few examples of different effects by disease aggressiveness.

In conclusion, our data provide additional evidence for a relation between T2D and PCa. Current investigations of a shared genetic background that could underlie this observed

association are still in their infancy but suggest that a genetic predisposition to T2D might also be associated with PCa risk. Future studies should further investigate the potential genetic factors that link these two common chronic diseases.

#### ACKNOWLEDGMENTS

Author affiliations: Program in Molecular and Genetic Epidemiology, Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts (Mitchell J. Machiela, Sara Lindström, David J. Hunter, Peter Kraft); Cancer Epidemiology Unit, Nuffield Department of Clinical Medicine, University of Oxford, Oxford, United Kingdom (Naomi E. Allen, Ruth Travis); Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California (Christopher A. Haiman, Brian E. Henderson, Daniel O. Stram); Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland (Demetrius Albanes, Sonja I. Berndt, Stephen Chanock); Navarre Public Health Institute, Pamplona, Spain (Aurelio Barricarte); National Institute for Public Health and the Environment, Bilthoven, the Netherlands (H. Bas Bueno-de-Mesquita); Department of Gastroenterology and Hepatology, University Medical Centre Utrecht, Utrecht, the Netherlands (H. Bas Bueno-de-Mesquita); Division of Aging, Brigham and Women's Hospital, Boston, Massachusetts (J. Michael Gaziano); Massachusetts Veterans Epidemiology Research and Information Center, VA Boston Healthcare System, Boston, Massachusetts (J. Michael Gaziano); Epidemiology Research Program, American Cancer Society, Atlanta, Georgia (Susan M. Gapstur, Eric J. Jacobs, Victoria L. Stevens); Department of Nutrition, Harvard School of Public Health, Boston, Massachusetts (Edward Giovannucci, Meir J. Stampfer, Walter C. Willett); Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts (Edward Giovannucci, Meir J. Stampfer, Walter C. Willett); University of Hawaii Cancer Center, Honolulu, Hawaii (Laurence N. Kolonel, Loic Le Marchand); Fondazione IRCCS [Istituto Di Ricovero e Cura a Carattere Scientifico], Istituto Nazionale dei Tumori, Milan, Italy (Vittorio Krogh); Channing Laboratory, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts (Jing Ma, Meir J. Stampfer); and Institute of Cancer Epidemiology, Danish Cancer Society, Copenhagen, Denmark (Anne Tjønneland).

This work was supported by the National Cancer Institute (cooperative agreements U01-CA98233-07 with David J. Hunter, U01-CA98710-06 with Susan M. Gapstur, U01-CA98216-06 with Elio Riboli and Rudolf Kaaks, and U01-CA98758-07 with Brian E. Henderson and the Intramural Research Program of the National Institutes of Health/National Cancer Institute, Division of Cancer Epidemiology and Genetics), as well as National Institutes of Health grants T32-CA09001 and T32-GM074897.

Conflict of interest: none declared.

## REFERENCES

1. Jemal A, Siegel R, Xu J, et al. Cancer statistics, 2010. *CA Cancer J Clin*. 2010;60(5):277–300.
2. Mokdad AH, Ford ES, Bowman BA, et al. Diabetes trends in the U.S.: 1990–1998. *Diabetes Care*. 2000;23(9):1278–1283.
3. Kasper JS, Giovannucci E. A meta-analysis of diabetes mellitus and the risk of prostate cancer. *Cancer Epidemiol Biomarkers Prev*. 2006;15(11):2056–2062.
4. Bonovas S, Filioussi K, Tsantes A. Diabetes mellitus and risk of prostate cancer: a meta-analysis. *Diabetologia*. 2004;47(6):1071–1078.
5. Kasper JS, Liu Y, Giovannucci E. Diabetes mellitus and risk of prostate cancer in the Health Professionals Follow-Up Study. *Int J Cancer*. 2009;124(6):1398–1403.
6. Kaaks R, Stattin P. Obesity, endogenous hormone metabolism, and prostate cancer risk: a conundrum of “highs” and “lows.” *Cancer Prev Res (Phila)*. 2010;3(3):259–262.
7. Giovannucci E, Michaud D. The role of obesity and related metabolic disturbances in cancers of the colon, prostate, and pancreas. *Gastroenterology*. 2007;132(6):2208–2225.
8. Winckler W, Weedon MN, Graham RR, et al. Evaluation of common variants in the six known maturity-onset diabetes of the young (MODY) genes for association with type 2 diabetes. *Diabetes*. 2007;56(3):685–693.
9. Gudmundsson J, Sulem P, Steinthorsdottir V, et al. Two variants on chromosome 17 confer prostate cancer risk, and the one in TCF2 protects against type 2 diabetes. *Nat Genet*. 2007;39(8):977–983.
10. Frayling TM. Genome-wide association studies provide new insights into type 2 diabetes aetiology. *Nat Rev Genet*. 2007;8(9):657–662.
11. Thomas G, Jacobs KB, Yeager M, et al. Multiple loci identified in a genome-wide association study of prostate cancer. *Nat Genet*. 2008;40(3):310–315.
12. Zeggini E, Scott LJ, Saxena R, et al. Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. Wellcome Trust Case Control Consortium. *Nat Genet*. 2008;40(5):638–645.
13. Eeles RA, Kote-Jarai Z, Al Olama AA, et al. Identification of seven new prostate cancer susceptibility loci through a genome-wide association study. *Nat Genet*. 2009;41(10):1116–1121.
14. Johnson AD, Handsaker RE, Pulit SL, et al. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics*. 2008;24(24):2938–2939.
15. Prokunina-Olsson L, Fu YP, Tang W, et al. Refining the prostate cancer genetic association within the JAZF1 gene on chromosome 7p15.2. *Cancer Epidemiol Biomarkers Prev*. 2010;19(5):1349–1355.
16. Stevens VL, Ahn J, Sun J, et al. HNF1B and JAZF1 genes, diabetes, and prostate cancer risk. *Prostate*. 2010;70(6):601–607.
17. Meyer TE, Boerwinkle E, Morrison AC, et al. Diabetes genes and prostate cancer in the Atherosclerosis Risk in Communities study. *Cancer Epidemiol Biomarkers Prev*. 2010;19(2):558–565.
18. Pierce BL, Ahsan H. Genetic susceptibility to type 2 diabetes is associated with reduced prostate cancer risk. *Hum Hered*. 2010;69(3):193–201.
19. Waters KM, Wilkens LR, Monroe KR, et al. No association of type 2 diabetes risk variants and prostate cancer risk: the Multiethnic Cohort and PAGE. *Cancer Epidemiol Biomarkers Prev*. 2011;20(9):1979–1981.
20. ATBC Cancer Prevention Study Group. The Alpha-Tocopherol, Beta-Carotene Lung Cancer Prevention Study: design, methods, participant characteristics, and compliance. *Ann Epidemiol*. 1994;4(1):1–10.
21. Calle EE, Rodriguez C, Jacobs EJ, et al. The American Cancer Society Cancer Prevention Study II Nutrition Cohort: rationale, study design, and baseline characteristics. *Cancer*. 2002;94(9):2490–2501.
22. Riboli E, Hunt KJ, Slimani N, et al. European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. *Public Health Nutr*. 2002;5(6B):1113–1124.
23. Severi G, Morris HA, MacInnis RJ, et al. Circulating insulin-like growth factor-1 and binding protein-3 and risk of prostate cancer. *Cancer Epidemiol Biomarkers Prev*. 2006;15(6):1137–1141.
24. Kolonel LN, Henderson BE, Hankin JH, et al. A multiethnic cohort in Hawaii and Los Angeles: baseline characteristics. *Am J Epidemiol*. 2000;151(4):346–357.
25. Prorok PC, Andriole GL, Bresalier RS, et al. Design of the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial. *Control Clin Trials*. 2000;21(6 suppl):p273S–309S.
26. Altshuler D, Hirschhorn JN, Klannemark M, et al. The common PPAR $\gamma$  Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. *Nat Genet*. 2000;26(1):76–80.
27. Gloyn AL, Weedon MN, Owen KR, et al. Large-scale association studies of variants in genes encoding the pancreatic beta-cell KATP channel subunits Kir6.2 (KCNJ11) and SUR1 (ABCC8) confirm that the KCNJ11 E23K variant is associated with type 2 diabetes. *Diabetes*. 2003;52(2):568–572.
28. Grant SF, Thorleifsson G, Reynisdottir I, et al. Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. *Nat Genet*. 2006;38(3):320–323.
29. Saxena R, Voight BF, Lyssenko V, et al. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science*. 2007;316(5829):1331–1336.
30. Sandhu MS, Weedon MN, Fawcett KA, et al. Common variants in WFS1 confer risk of type 2 diabetes. *Nat Genet*. 2007;39(8):951–953.
31. Sladek R, Rocheleau G, Rung J, et al. A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature*. 2007;445(7130):881–885.
32. Yasuda K, Miyake K, Horikawa Y, et al. Variants in KCNQ1 are associated with susceptibility to type 2 diabetes mellitus. *Nat Genet*. 2008;40(9):1092–1097.
33. Rung J, Cauchi S, Albrechtsen A, et al. Genetic variant near IRS1 is associated with type 2 diabetes, insulin resistance and hyperinsulinemia. *Nat Genet*. 2009;41(10):1110–1115.
34. Prokopenko I, Langenberg C, Florez JC, et al. Variants in MTNR1B influence fasting glucose levels. *Nat Genet*. 2009;41(1):77–81.
35. Dupuis J, Langenberg C, Prokopenko I, et al. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat Genet*. 2010;42(2):105–116.
36. Voight BF, Scott LJ, Steinthorsdottir V, et al. Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat Genet*. 2010;42(7):579–589.
37. Schumacher FR, Berndt SI, Siddiq A, et al. Genome-wide association study identifies new prostate cancer susceptibility loci. *Hum Mol Genet*. 2011;20(19):3867–3875.
38. Li Y, Willer CJ, Ding J, et al. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet Epidemiol*. 2010;34(8):816–834.
39. Westfall PH, Zaykin KV, Young SS. Multiple tests for genetic effects in association studies. In: Looney SW, ed.

- Methods in Molecular Biology*. Totowa, NJ: Humana Press, Inc.; 2002:143–168.
40. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81(3):559–575.
  41. Kwee LC, Liu D, Lin X, et al. A powerful and flexible multilocus association test for quantitative traits. *Am J Hum Genet*. 2008;82(2):386–397.
  42. Wu MC, Kraft P, Epstein MP, et al. Powerful SNP-set analysis for case-control genome-wide association studies. *Am J Hum Genet*. 2010;86(6):929–942.
  43. Lindstrom S, Schumacher F, Siddiq A, et al. Characterizing associations and SNP-environment interactions for GWAS-identified prostate cancer risk markers—results from BPC3. *PLoS One*. 2011;6(2):pe17142. (doi:10.1371/journal.pone.0017142).
  44. Baron RM, Kenny DA. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *J Pers Soc Psychol*. 1986;51(6):1173–1182.
  45. Vanderweele TJ, Vansteelandt S. Odds ratios for mediation analysis for a dichotomous outcome. *Am J Epidemiol*. 2010;172(12):1339–1348.
  46. Garup N, Sparsø T, Hansen T. Physiologic characterization of type 2 diabetes-related loci. *Curr Diab Rep*. 2010; 10(6):485–497.
  47. Artavanis-Tsakonas S, Rand MD, Lake RJ. Notch signaling: cell fate control and signal integration in development. *Science*. 1999;284(5415):770–776.
  48. Thomas G, Jacobs KB, Kraft P, et al. A multistage genome-wide association study in breast cancer identifies two new risk alleles at 1p11.2 and 14q24.1 (RAD51L1). *Nat Genet*. 2009;41(5):579–584.
  49. Fu YP, Edvardsen H, Kaushiva A, et al. NOTCH2 in breast cancer: association of SNP rs11249433 with gene expression in ER-positive breast tumors without TP53 mutations. *Mol Cancer*. 2010;9:p113. (doi:10.1186/1476-4598-9-113).
  50. Orr B, Grace OC, Vanpoucke G, et al. A role for notch signaling in stromal survival and differentiation during prostate development. *Endocrinology*. 2009;150(1): 463–472.
  51. Midthjell K, Holmen J, Bjørndal A, et al. Is questionnaire information valid in the study of a chronic disease such as diabetes? The Nord-Trøndelag diabetes study. *J Epidemiol Community Health*. 1992;46(5):537–542.
  52. Okura Y, Urban LH, Mahoney DW, et al. Agreement between self-report questionnaires and medical record data was substantial for diabetes, hypertension, myocardial infarction and stroke but not for heart failure. *J Clin Epidemiol*. 2004;57(10):1096–1103.
  53. Centers for Disease Control and Prevention. *National Diabetes Fact Sheet: General Information and National Estimates on Diabetes in the United States, 2007*. Atlanta, GA: Centers for Disease Control and Prevention; 2008.
  54. Waters KM, Henderson BE, Stram DO, et al. Association of diabetes with prostate cancer risk in the Multiethnic Cohort. *Am J Epidemiol*. 2009;169(8):937–945.