Quantifying the Contributions of Environmental Factors to Prostate Cancer and Detecting Risk-Related Diet Metrics and Racial Disparities

$\circledS$SAGE

Wensheng Zhang[1] and Kun Zhang[1,2]

[1]Bioinformatics Core of Xavier NIH RCMI Center of Cancer Research, Xavier University of Louisiana, New Orleans, LA, USA. [2]Department of Computer Science, Xavier University of Louisiana, New Orleans, LA, USA.

ABSTRACT: The relevance of nongenetic factors to prostate cancer (PCa) has been elusive. We aimed to quantify the contributions of environmental factors to PCa and identify risk-related diet metrics and relevant racial disparities. We performed a unique analysis of the Diet History Questionnaire data of 41 830 European Americans (EAs) and 1282 African Americans (AAs) in the PLCO project. The independent variables in the regression models consisted of age at trial entry, race, family history of prostate cancer (PCa-fh), diabetes history, body mass index (BMI), lifestyle (smoking and coffee consumption), marital status, and a specific nutrient/food factor (X). $P < .05$ and a 95% confidence interval excluding zero were adopted as the criteria for determining a significant difference (effect). We established a priority ranking among PCa risk-related genetic and environmental factors according to the deviances explained by them in the multivariate Cox-PH regression analysis: age > PCa-fh > diabetes ≥ race > lifestyle ≥marital-status ≥BMI > X. We confirmed previous studies showing that (1) high protein and saturated fat levels in diet were related to increased PCa risk, (2) high-level supplementary selenium intake was harmful rather than beneficial for preventing PCa, and (3) supplementary vitamin B6 was beneficial for preventing benign PCa. We obtained the following novel findings: high-level organ meat intake was an independent predictor for increased aggressive PCa risk; supplementary iron, copper and magnesium increased benign PCa risk; and the AA diet was "healthy" in terms of the relatively lower protein and fat levels and was "unhealthy" in that it more commonly contained organ meat. In conclusion, we established a priority ranking among the contributing factors for PCa and identified several risk-related diet metrics and the racial disparities. Our findings suggested some new approaches to prevent PCa such as restriction of organ meat intake and supplementary microminerals.

KEYWORDS: Environmental factors, prostate cancer, racial disparities, diet metrics, nutrition

Introduction

Aside from non-melanoma skin cancer, prostate cancer (PCa) is the most frequently occurring cancer in American males, but its prevalence in Asian countries is very low. Recent genetic studies demonstrated that PCa susceptibility is a polygenic trait with the genomic heritability ranging from 0.25 to 0.78,[1,2] while pathogenic germline variants in some cancer genes such as BRCA1, BRCA2, and HOXB13 can be detected in a small fraction of men with a personal history of the disease.[3] The contribution of nongenetic, that is, environmental, factors to this trait is strongly suggested by the fact that The Asian immigrants in Western countries have higher incidence of prostate cancer compared to those in their countries of birth.[4] Potential environmental risk factors for PCa include diet metrics such as nutrient intake levels, lifestyle patterns such as those characterized by smoking and coffee consumption, health status related to other diseases, and actual marital status.[5,6]

The implications of diet and lifestyle characteristics for PCa epidemiology and prevention have not been fully determined. In particular, previous studies yielded inconsistent results on the effects of high- or low-level intake of macronutrients, including carbohydrates, protein, fat, calcium and phosphorous, on the PCa risk. For example, a study of 142 251 men in the European Prospective Investigation into Cancer and Nutrition supported the hypothesis that high-level intake of protein or calcium from dairy products may increase the risk for prostate cancer.[7] However, such associations were not presented in another study on an American cohort, which included 11 014 participants.[8] Meat is another main dietary protein source, especially in Western society. A positive association between red meat (pork, beef, and lamb) intake and the prostate cancer risk was suggested by several studies but was not confirmed by others.[9-12] Similarly, uncertainty widely exists in the observed relationships between the PCa risk and the intake of other macronutrients and major food ingredients.[6]

The liability of dietary and lifestyle data, which are usually derived from questionnaires, is subject to many uncertain factors. In particular, a dietary record collected close to the time of cancer diagnosis or at the end of the follow-up period for a "control" may lack relevance to the actual environment for cancer initiation and progression because there is a period of ~10 years between the genesis of initial tumorous cells and the final formation of a tumor that can be detected by conventional

methods.[13] This may be a major reason for the inconsistency among the reported relationships between cancer susceptibility and diet/lifestyle factors across studies. The situation is further complicated by the interweaving of genetic factors and environmental factors and the interactions between environmental factors themselves. For example, diabetes and obesity may be considered environmental factors for prostate cancer. Meanwhile, they may share some common genetic risk factors with cancer.[14,15] More importantly, diabetes and obesity statuses also influence a subject's food and lifestyle choices. In such a context, carefully selected data preprocessing procedures and statistical models are crucial for uncovering relationships hidden within "noisy" data.

In this study, we aimed to quantify the contributions of non-genetic risk factors to prostate cancer, identify risk-associated dietary variables, and assess correlations between dietary variables and blood prostate-specific antigen (PSA) level, which is a commonly used biomarker for detecting PCa. We also investigated racial disparities in those dietary risk metrics, which are important for understanding and alleviating the substantial inequality in PCa prevalence across races.[16] We pursued these goals by analyzing the data collected by the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial (PLCO) project.[17] This data has been used in many studies, such as Cross et al.,[12] Shikany et al.,[18] Richman et al.[19] but are worthy of further analysis. Our study differs from others in the data refining/preprocessing procedure and statistical models used (see the Methods and Discussion sections). For the associations between diet metrics and PCa risk, our general hypothesis was that both confirmative results, such as the increase of PCa risk due to high-level saturated fat and red meat intake, and novel findings may be achieved via the performed analyses.

Material and Methods

Datasets

The Prostate Person dataset in the file "pros_data_nov18_d070819.csv" and Diet History Questionnaire (DHQ) dataset in the file "dhq_data_nov18_d070819.csv" were downloaded from the PLCO database upon approval from the National Institutes of Health (NIH). The DHQ data was collected by the researchers blinded to the clinical outcomes of the PLCO's participants. In this study, the 2 datasets were first merged on subjects' trial IDs. Then, records without complete information for the variables considered in this study were filtered out. The subjects for whom the time from the date of trial entry, i.e., the date when a participant was randomized into the screening or control arm, to the DHQ entry date was longer than 6 years and those with a personal PCa history preceding the date of trial entry were also excluded. After this refinement, the working dataset (WD) included the information of 41 830 Non-Hispanic European American (EA) males and 1282 Non-Hispanic African Americans (AA) males. In Table 1, we summarize the distributions of subjects in the WD regarding

the factors studied in our basic statistical model (see the following Statistical methods section).

Diet metrics

Two sets of diet metrics (variables) were tested in the statistical analysis. The first set included 29 metrics for dietary nutrients (dt_x, x is protein or others), 21 supplementary nutrients (sup_y_dhq, y is selenium or others) and the ratio of dt_calcium and dt_phosphorus. The dt_x and sup_y metrics were calculated from the questionnaire responses by DietCalc software, which considered food frequency, serving size, and other responses. The second set included 30 metrics for food ingredients (p_z, z is milk or others). The p_z metrics were calculated in the same manner as the nutrient variables using amounts based on the US Department of Agriculture's pyramid servings, which are stored in the same gender- and serving size-specific database as the nutrients.

Data preprocessing

We modified diet metrics using a 2-step procedure. First, the metrics were adjusted for age using a simple linear model. Second, the adjusted metrics were normalized with total calorie intake as the scale factor. As such, the normalized dt_x (or supp_y, or p_z) was the intake of x dietary nutrient (or y supplementary nutrient, or z dietary food ingredient) per Kcal total energy. Each subject's PSA was measured in multiple screenings. For a noncancer subject, we took the average of the PSA levels over the screenings as the representative value (RV). For a case subject, we took the pre-diagnosis PSA level, that is, the measure obtained at the screening immediately preceding the PCa diagnosis, as the RV.

Statistical methods

A series of Cox proportional hazards regression analyses, where the PCa indicator (1 or 0) and disease-free survival time from the date of trial entry to the end of the follow-up were dependent variables, were conducted to assess the associations between explanatory variables, such as dietary (dt_) protein, and the disease, that is, prostate cancer. The analysis procedure proceeded as follows. First, a univariate model was used to scan the Basic Questionnaire (BQ) variables (except for caffeine consumption), which were contained in the "Prostate person" dataset, to detect factors associated with PCa. Second, the effects of the significant BQ factors determined in the last step were reassessed using a multivariate model, which is called the basic model (M1) hereafter. Finally, the association between a diet metric and PCa was tested using another multivariate model (M2) in which all the items in M1, alongside the diet metric itself, were included as covariates. In the M2 analysis, the age-adjusted and normalized diet metrics were discretized into 4 levels for dietary nutrients and common food ingredients or 2

Table 1. Summary of samples used in the analysis of the risk for aggressive and benign cancer.

	NON CANCER		AGGRESSIVE CANCER				BENIGN CANCER			
	EA	AA	EA	% EA[a]	AA	% AA[b]	EA	% EA[c]	AA	% AA[d]
Age level 0	14110	474	588	3.8	33	6.1	791	5.1	34	6.3
Age level 1	12961	387	893	6	41	8.6	1149	7.7	46	9.7
Age level 2	9540	278	680	6.1	32	9.4	936	8.4	30	8.8
Age level 3	5225	143	353	5.9	16	8.8	413	6.9	23	12.6
Pros_fh-	38826	1183	2219	5.0	110	7.8	2913	6.6	124	8.8
pros_fh+	3010	99	295	8.0	12	10	376	10.2	9	7.5
Smoke- & Coffee -	14009	564	951	5.9	57	8.3	1223	7.6	62	9.1
Smoke- & Coffee+	21404	340	1271	5.2	33	8.1	1662	6.8	36	8.8
Smoke+ &Coffee-	1423	192	58	3.7	15	6.8	97	6.1	15	6.8
Smoke+ & Coffee+	5000	186	234	4.2	17	7.6	307	5.5	20	9.0
Diab-	38595	1056	2369	5.4	105	8.2	3128	7.1	121	9.4
Diab+	3241	226	145	4.1	17	6.7	161	4.5	12	4.7
BMI 1	10629	329	680	5.5	32	8	962	7.8	37	9.3
BMI 2	21502	566	1308	5.3	60	8.7	1708	7.0	64	9.3
BMI 3	9705	387	526	4.8	30	6.7	619	5.7	32	7.1
Marital status-	35745	868	2239	5.5	86	8.2	2902	7.1	95	9.1
Marital status+	6091	414	275	4.1	36	7.4	387	5.7	38	7.8

[a,b]The percentages of aggressive EA or AA PCa cases among all EA or AA samples. [c,d]The percentages of benign EA or AA PCa cases among all EA or AA samples.

levels for supplementary nutrients and uncommon food ingredients such as organ meat. Four-level discretization was established using the 25%, 50%, and 75% quantiles of the original measures as the cutoffs. Two-level discretization was established using the mean as the cutoff.

$P < 0.05$ and a 95% confidence interval (95% CI) that did not contain zero in the terms of effect coefficient, that is, did not contain 1.0 in hazards ratio or relative risk (RR) for prostate cancer, were adopted as the criteria for determining the significance of a difference (effect). For a categorical explanatory variable, RR indicated the risk ratio between a test group such as "diabetes+" and the baseline group such as "diabetes-". The null hypothesis to be tested was $H_0 : \alpha = 0$, where α was the coefficient for the test group given the coefficient of the baseline group being zero. For a quantitative explanatory variable, RR indicated the change in risk if the metric increased by one unit. The null hypothesis to be tested was $H_0 : \beta = 0$, where β was the coefficient for the variable. The contribution of a variable (x) to prostate cancer was quantified with the deviance explained by x, which was 2 times the log-likelihood ratio of the full model that included x compared to the reduced model that did not include x.

The correlations between PSA and diet metrics were calculated using Pearson's method and were tested using a t test. Preceding this step, individuals' PSA values were adjusted for the trial entry ages. Two methods were used to compare AAs and EAs in the intake of nutrients and foods. For the metrics of dietary nutrients and common foods, a t test was employed. For a supplementary nutrient or uncommon food ingredient, persons were partitioned into a high-intake group and a low-intake group with the mean intake as the cutoff, and then a chi-square test was performed.

With regard to the multiple testing for the associations between the focused diet metrics and PCa risk, the correlations between diet metrics and PSA levels and the differences of diet metrics between EA males and AA males, the false discover rate (FDR) corresponding to an ordinary p-value was estimated by the Benjamini–Hochberg procedure.

Software application

All analyses and graphics were completed by using the relevant functions in the R packages, including "stats," "survival" and "ltm," as well as our lab-owned R codes.

Results

Basic model analysis for the PCa risk

The basic model M1 (see the Statistical methods section) included 7 categorical explanatory variables: age level, PCa-fh, race, diabetes history, body mass index (BMI, a measure of body fat based on height and weight) class, lifestyle (smoking and coffee consumption) and marital status. Except for coffee consumption, which was coded on the DHQ entry, the other variables were coded on the information collected at the dates of trial entry. Age level was coded with "1","2", "3", or "4" if a subject's age fell within the interval 45 to 59, 60 to 64, 65 to 69 or over 70, respectively. PCa-fh was coded with "+" or "-", depending on the existence of PCa family history in first-degree relatives. Diabetes history was coded with "+" if a subject had diabetes and otherwise was coded with "-." BMI class was coded "1", "2" or "3" if the quantity fell within the interval 0 to 25, 25 to 30 or 30+, respectively. Smoking was coded with "+" if the subject's cigarette smoking status was "current cigarette smoker" or he had a smoking history over 5 years and the time from the date when he stopped smoking to the date of trial entry was less than 5 years and otherwise coded with "-". Coffee consumption was coded with "+" if a subject's consumption of caffeine was ≥0.1 mg/day and otherwise was coded with "-". Lifestyle was defined as the combination of smoking and coffee consumption. Marital status was coded with "+" if a subject was married or living as married and otherwise was coded with "-".

The analysis results of the M1 model are presented in Table 2, which was partitioned into modules **I**, **II**, and **III**. In **I**, the information of all prostate cancer cases (alongside noncancer subjects) was used. In **II** and **III**, the information of aggressive cases (ie, Gleason score ≥7 and/or T-stage III-IV) and benign cases (i.e., Gleason score <7 and T-stage I-II) was used. The baselines for the 7 variables were age level 1, pros_fh-, race EA, smoke-&coffee-, diabetes-, BMI 1 and marital status-, respectively. Age was the top contributory factor to both aggressive and benign cancer, but the 95% CIs for the relative risk of age levels 2 and 3 partially overlapped, indicating that the association is mainly due to the difference in PCa risk between the 45 and 59 age group and the others. The second important risk factor was pros_fh, with the explained deviance being approximately half of the deviance explained by age. The third significant factor was race for aggressive cancer or diabetes for benign cancer. Later, lifestyle and marital status had similar contributions to both aggressive cancer and benign cancer, with the explained deviance approximating half of the deviance explained by race or diabetes. BMI significantly contributed to benign cancer but not aggressive cancer. Summarizing these results, we established a priority ranking of the variables considered in M1 according to the explained deviances: age >PCa-fh > diabetes ≥ race > lifestyle (smoking and coffee consumption) ≥ marital status ≥BMI. We also noted that, in reference to those reported by previous publications, the specific results from this analysis included confirmative components such as the positive association between cigarette smoking and PCa and controversial components such as the negative association between BMI and cancer (see Discussion section).

PCa risk–related diet metrics

The associations between the PCa risk and diet metrics were assessed by the model M2 as described in the Materials and Methods section. The analysis was first performed in the context that the discretized intake levels of a nutrient or food ingredient were considered quantitative amounts (Figure 1, Table 3). Nine, six, and five diet metrics were determined to be significantly associated with the PCa risk when the information of all cancer cases, aggressive cancer cases or benign cancer cases was used, respectively. The results confirmed the previously reported positive associations between the PCa risk and the intake levels of saturated fat,[20,21] red meat,[9] alcohol,[22,23] and supplementary selenium,[24] and the negative association between benign cancer and vitamin B6.[25] The novel findings included the positive association between benign cancer and supplementary iron/magnesium/copper, and the positive association between aggressive cancer and organ meat. In terms of the explained deviances, the priority of any diet metric for PCa risk was behind all other exploratory variables in the analysis model (see Table 2, Figure 1, and Supplemental Figures S1 and S2).

Then, the analysis was repeated for the aforementioned PCa-related dietary nutrients and common food ingredients in the alternative context that the discretized intake levels were considered categorical metrics. The sequentially increased intake levels 1, 2, 3, and 4 were replaced with category identifiers Q1, A2, Q3, and Q4, and Q1 was treated as the baseline in statistical tests. The analysis results are presented in Table 4. From them, we found some meaningful patterns regarding the relationship between a nutrient/ingredient and the PCa risk. For example, in the scenario of aggressive cancer and saturated fat, a single-step pattern of risk increases was apparent, as indicated by the similar RR values and 95% CIs for the contrasts Q2 versus Q1, Q3 versus Q1 and Q4 versus Q1. The estimated RR value of 1.11 to 1.14 denoted an 11% to 14% increase in PCa risk.

It is worth noting that the detected associations between PCa risk and diet metrics were moderate with the FDR values

Table 2. Results from the basic model (M1) analysis for the associations between PCa occurrence and risk factors.[a]

	RR	95% CI	*P*	DEV[B]
I All cancer				
age level 2	1.46	1.36-1.56	$<1 \times 10^{-16}$	205.01
age level 3	1.60	1.49-1.71	$<1 \times 10^{-16}$	–
age level 4	1.51	1.38-1.64	$<1 \times 10^{-16}$	–
pros_fh+	1.58	1.46-1.71	$<1 \times 10^{-16}$	114.75
race AA	1.67	1.47-1.89	2×10^{-15}	54.33
smoke-&coffee+	0.92	0.88-0.98	.0044	20.92
smoke+ & coffee-	0.82	0.71-0.96	.012	–
smoke+ & coffee+	0.84	0.76-0.92	1×10^{-4}	–
diabetes+	0.71	0.63-0.79	2×10^{-9}	39.60
BMI 2	0.93	0.88-0.99	.016	17.18
BMI 3	0.86	0.79-0.92	4×10^{-5}	–
Marital status+	0.84	0.78-0.91	1×10^{-5}	19.65
II Aggressive cancer:				
age level 2	1.47	1.33-1.63	1×10^{-13}	87.10
age level 3	1.56	1.40-1.74	1×10^{-15}	–
age level 4	1.56	1.37-1.76	3×10^{-11}	–
pros_fh+	1.65	1.47-1.86	1×10^{-16}	60.15
Race AA	1.88	1.56-2.26	2×10^{-11}	37.71
smoke-&coffee+	0.92	0.85-1.00	.042	11.83
smoke+& coffee-	0.74	0.59-0.94	.015	–
Smoke+ & coffee+	0.84	0.73-0.97	.014	–
Diabetes+	0.78	0.66-0.91	.002	10.09
BMI 2	0.97	0.89-1.06	.48	1.24
BMI 3	0.94	0.84-1.05	.27	–
Marital status+	0.81	0.72-0.91	6×10^{-4}	12.47
III Benign cancer:				
age level 2	1.47	1.35-1.61	$<1 \times 10^{-16}$	122.10
age level 3	1.63	1.49-1.80	$<1 \times 10^{-16}$	–
age level 4	1.41	1.25-1.59	1×10^{-8}	–
pros_fh+	1.58	1.42-1.75	$<1 \times 10^{-16}$	62.65
Race AA	1.54	1.29-1.84	2×10^{-6}	20.43
smoke- & coffee+	0.93	0.86-1.00	.038	12.42

(Continued)

Table 2. (Continued)

	RR	95% CI	*P*	DEV[B]
smoke+ & coffee-	0.87	0.72-1.06	.17	–
smoke+& coffee+	0.82	0.72-0.92	.0012	–
diabetes+	0.65	0.55-0.75	3×10^{-8}	34.94
BMI 2	0.90	0.83-0.97	.0053	23.56
BMI 3	0.78	0.70-0.87	2×10^{-6}	–
Marital status+	0.85	0.77-0.95	.0024	9.55

[a]Age level 1, pros_fh-, race EA, smoke-&coffee-, diabetes-, BMI 1 and marital status- were considered as the baselines for the seven categorical risk factors. [b]The deviance explained by the corresponding factor.

≥0.36 (Table 3). This implies that the findings warrant further validation. However, in interpreting a confirmative (to previous studies) observation, there is no strong rationale to use FDR or a multiple testing adjusted *P*-value.[26]

Racial disparities in diet and nutrition

The study of racial disparity in diet and nutrition was focused on PCa-related diet metrics (see Table 3). Two sets of analyses were performed on cancer samples and noncancer samples. As shown in Table 5, for all 15 diet metrics, the differences between the AA and EA noncancer groups were significant ($P < .05$). The results from the analysis of cancer samples were consistent with those of noncancer samples in terms of the estimated fold changes. However, the differences between the AA cancer group and the EA cancer group in dietary poultry and supplementary vitamin B6 were not significant ($P > .05$). This may be due to the small sample size (N = 265) of the AA cancer group. Only in dietary carbohydrates and organ meat did AA have higher intake levels than EA, regardless of whether the males were diagnosed with prostate cancer during the follow-up. In other nutrients/ingredients, the intake levels of AA were consistently lower than those of EA. In general, the racial disparity was moderate in the macro nutrients but was substantial in food ingredients and supplementary micronutrients. For example, in the scenario of cancer and protein, the log2-fold change (FC) was –0.114, indicating that AA (average) intake was 92% ($100 \times 2^{-0.114}$) of EA intake; in the scenario of cancer and red meat, the log2 (FC) was –0.411, indicating that AA intake was 75% ($100 \times 2^{-0.411}$) of EA intake.

Correlations between PSA and diet metrics

The association between each of the 80 diet metrics and PSA level was assessed by correlation analysis, which was first performed in cancer samples and then replicated in noncancer samples. The detected significant correlations, in which 28 metrics were involved, are summarized in Table 6. These correlations were moderate, with absolute r values ranging from .012 to .044. For a specific of diet metric, the correlations in cancer samples and noncancer samples were largely consistent. An exception was dietary saturated fat, for which the correlation was significant ($P = .02$) and relatively substantial ($r = .04$) in cancer samples but not in noncancer samples. For dietary protein and red meat, the negative correlations deviated from the conventional implication of their positive association with PCa risk. For dietary dairy and milk, the positive correlations were well fit with their positive association with cancer. The latter pattern was also true for dietary saturated fat and supplementary selenium.

Discussion

We performed a unique analysis of the PLCO Diet History Questionnaire data (along with the Prostate Person dataset). The achieved results are summarized and discussed here.

First, we established a priority ranking, or importance ranking, among PCa risk-related genetic and environmental metrics according to the deviances explained by them in the multivariate Cox-PH regression analysis; that is: age >PCa-fh (family history of prostate cancer) > diabetes ≥ race > lifestyle (smoking and coffee consumption) ≥ marital status ≥BMI > X, in which X represented a specific diet nutrient/ingredient metric. Age at the date of trial entry and PCa-fh were determined to be in the top places of the rank without ambiguity, reflecting the well-known knowledge that PCa is more likely to develop in older men and those with a family history of the disease.[6] In the third to fourth ranks were diabetes history and race. Diabetes is a hereditary disease, and its relevance to PCa has been reported in the literature.[27,28] Regarding this relationship, a heuristics hypothesis is that the two diseases may share common genetic risk factors, but supporting evidence for that has been weak.[29] An alternative hypothesis is that diabetes can change the internal environment for the initiation, progression or diagnosis of prostate cancer or influence cancer-related lifestyle sport and diet compositions.[6] Our results supported the latter hypothesis because the association between diabetes and PCa occurrence
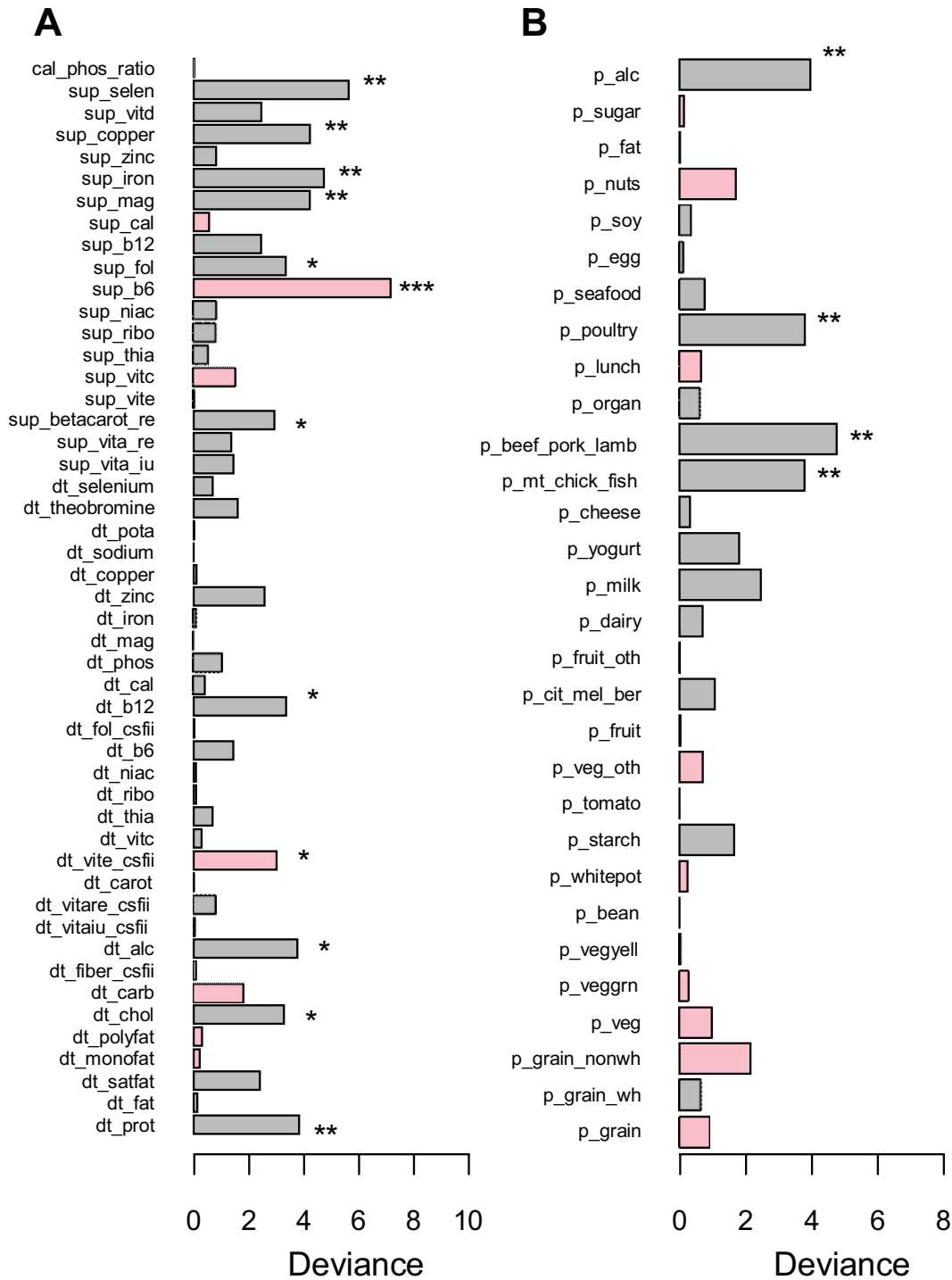
Figure 1. Contributions of diet metrics to prostate cancer risk, estimated using the information of all cancer and noncancer samples. The bar color, that is, gray or pink, indicates the negative or positive association between the corresponding diet metric and cancer risk. *, **, and *** denote significance levels with *P*-value less than .1, .05, and .01.

was negative rather than positive. The priority of race may be underestimated due to the inclusion of PCa-fh in the analysis model. In the last 3 ranks were the lifestyle defined by smoking and coffee consumption, marital status and BMI. The negative effects of cigarette smoking and high-level consumption of coffee on the PCa risk or diagnosis confirmed previous studies on this issue.[30,31] The negative association of "married" living status

with the PCa risk confirmed the finding in Ref[5] and may share a common mechanism with the connection between the frequency of ejaculation throughout adult life and prostate cancer occurrence, which was reported in Ref.[32] The negative association of a high BMI with PCa was somewhat conflicting with the conventional conception about the relationship between obesity and prostate cancer.[33,34]

Table 3. Association between the PCa risk and diet metrics.[a]

	CANCER	RR	95% CI	*P*-VALUE	FDR
dt_prot	whole	1.023	1.000-1.047	.05	0.36
sup_b6	whole	0.894	0.823-0.971	.01	0.36
sup_mag	whole	1.056	1.003-1.110	.04	0.36
sup_iron	whole	1.058	1.006-1.113	.03	0.36
sup_copper	whole	1.055	1.003-1.110	.04	0.36
sup_selen	whole	1.129	1.023-1.246	.02	0.36
p_beef_pork_lamb	whole	1.025	1.003-1.050	.03	0.36
p_poultry	whole	1.024	1.000-1.046	.05	0.36
p_alc	whole	1.024	1.001-1.047	.05	0.36
dt_satfat	aggressive	1.042	1.006-1.079	.04	0.52
dt_carb	aggressive	0.963	0.930-0.998	.04	0.52
sup_iron	aggressive	1.082	1.001-1.169	.05	0.52
p_beef_pork_lamb	aggressive	1.036	1.000-1.073	.05	0.52
p_mt_chick_fish	aggressive	1.037	1.002-1.074	.04	0.52
p_organ	aggressive	1.125	1.031-1.227	.01	0.52
dt_vite__dhq	benign	0.970	0.942-1.000	.05	0.50
dt_alc	benign	1.031	1.001-1.047	.05	0.50
sup_b6	benign	0.872	0.780-0.974	.01	0.50
sup_selen	benign	1.179	1.037-1.341	.01	0.50
p_milk	benign	1.031	1.000-1.062	.05	0.50

[a]Identifiers of diet metrics, that is, nutrients and food ingredients, are consistent with those presented in the PLCO DHQ dataset. dt_: dietary; sup_: supplementary; p_: pyramid serving; prot: protein; alc: alcohol; b6: vitamin B6; mag: magnesium; selen: selenium; satfat: saturated fat; carb: carbohydrates; vite: vitamin E.

Table 4. Differences in PCa risk between quantile levels in the intake of nutrients and food ingredients.[a]

	CANCER	Q2 VERSUS Q1		Q3 VERSUS Q1		Q4 VERSUS Q1	
		RR	95% CI	RR	95% CI	RR	95% CI
dt_prot	whole	1.007	0.937-1.081	1.028	0.957-1.105	1.072	0.997-1.152
p_alc	whole	0.98	0.908-1.051	1.088	1.013-1.168	1.043	0.970-1.121
p_beef_pork_lamb	whole	1.003	0.933-1.077	1.036	0.964-1.112	1.078	1.003-1.158
p_poultry	whole	1.062	0.989-1.142	1.053	0.980-1.132	1.082	1.008-1.163
dt_satfat	aggressive	1.110	0.995-1.237	1.141	1.022-1.273	1.137	1.018-1.270
dt_carb	aggressive	1.126	1.011-1.254	1.035	0.928-1.156	0.908	0.810-1.019
p_mt_chick_fish	aggressive	1.054	0.945-1.177	1.101	0.987-1.228	1.114	0.998-1.243
p_beef_pork_lamb	aggressove	1.009	0.905-1.126	1.057	0.948-1.179	1.107	0.992-1.236
dt_vite_csfii_dhq	Benign	1.000	0.914-1.101	0.923	0.839-1.016	0.930	0.845-1.022
dt_alc	Benign	0.992	0.899-1.094	1.111	1.010-1.223	1.068	0.970-1.176
p_milk	Benign	1.070	0.972-1.178	1.106	1.005-1.218	1.068	0.970-1.176

[a]The identifier annotations of nutrients and food ingredients are the same as those in Table 3.

Table 5. Racial disparities in nutrient and food intake.[a]

	NON CANCER			CANCER		
	LOG2(FC)	*P*-VALUE	FDR	LOG2(FC)	*P*-VALUE	FDR
dt_prot	−0.144	<.0001	<0.0001	−0.148	<.0001	<0.0001
dt_alc	−1.333	<.0001	<0.0001	−1.284	<.0001	<0.0001
dt_satfat	−0.093	<.0001	<0.0001	−0.117	.0001	0.0002
dt_carb	0.067	<.0001	<0.0001	0.079	<.0001	<0.0001
dt_monofat	−0.041	.001	0.001	−0.038	.1746	0.19
p_beef_pork_lamb	−0.411	<.0001	<0.0001	−0.355	<.0001	<0.0001
p_poultry	−0.088	.0182	0.19	−0.146	.0713	0.09
p_mt_chick_fish	−0.105	<.0001	<0.0001	−0.101	.0353	0.05
p_organ	0.293	<.0001	<0.0001	0.344	.018	0.03
p_milk	−0.577	<.0001	<0.0001	−0.719	<.0001	<0.0001
sup_b6	−0.296	.0252	0.02	0.212	.4913	0.49
sup_mag	−0.522	<.0001	<0.0001	−0.337	.0049	0.009
sup_iron	−0.417	<.0001	<0.0001	−0.281	.0147	0.02
sup_copper	−0.522	<.0001	<0.0001	−0.337	.0049	0.009
sup_selen	−0.779	.0002	0.0003	−0.742	.0003	0.12

[a]The identifier annotations of nutrients and food ingredients are the same as those in Table 3.

Second, we identified over a dozen diet metrics that were associated with the PCa risk. The confirmative observations included the following: a saturated fat diet was related to an increased risk for aggressive PCa,[21,35] and a high-level intake of supplementary selenium may be harmful rather than beneficial for preventing PCa,[24] and supplementary vitamin B6 was beneficial for preventing benign PCa. The novel findings included the following: organ meat intake, as an independent predictor rather than as a contributing component of a PCa-related dietary pattern composed of the intakes of many foods,[36] was positively associated with aggressive PCa risk; and supplementary iron, copper and magnesium increased benign PCa risk. Our results also demonstrated that PCa risk was associated with the intake levels of protein and a few major animal protein ingredients, such as red meat, poultry and milk. We noted that in an early publication that was based on the analysis of the midway data of the PLCO project,[12] Cross and her colleagues reported that very well-performed meat was positively associated with PCa risk, but they did not observe any association between PCa risk and red/white meat intake. The inconsistency between Cross et al's result and our result may be attributed to the differences in the data preprocessing procedure and statistical models. For example, age, not clearly specified in Ref,[12] was considered as the underlying time metric in the Cox-PH regression model used in Cross et al's work. However,

in our model, the time metric was disease-free survival years from the date of trial entry to the end of the follow-up, and the age at the date of trial entry was considered a covariate. The associations observed here between PCa risk and the intakes of several micronutrients (i.e., iron, copper, magnesium and vitamin B6) may hold significant implications for cancer prevention. The reason is that controlling the intake of these supplementary nutrients is much more practical and convenient than modifying genetic factors and other environmental factors. Moreover, the potential of vitamin B6 as a cancer risk reduction agent was also reported in a previous study of gastrointestinal cancer.[37] An issue worthy of further investigation is that, among iron, copper and magnesium, which one was actually responsible for their association with PCa risk. Given that microminerals are usually supplemented by taking multimineral product(s), we could not solve the uncertainty using the PLCO questionnaire data.

Third, regarding the PCa risk-related diet metrics, we pinpointed a few racial disparities. Based on the observed differences, we conceived that the AA diet was "healthy" in terms of the relatively lower protein and fat levels and was "unhealthy" in that it more commonly contained organ meat. However, for deciphering racial disparity in PCa occurrence, the implication of these results should be considered to be limited. This postulation can be further scrutinized in the following manner.

Table 6. Correlations between diet metrics and PSA level.[a]

	NON CANCER			CANCER		
	R	*P*-VALUE	FDR	*R*	*P*-VALUE	FDR
dt_prot	−0.015	.03	0.11	−0.012	.48	0.62
dt_satfat	−0.002	.82	0.84	.04	.022	0.20
dt_carb	.024	.0004	0.005	.009	.63	0.75
dt_niac	−0.023	.0005	0.005	−0.051	.004	0.20
dt_cal	.016	.014	0.07	.021	.23	0.44
dt_zinc	−0.016	.018	0.08	.009	.60	0.75
dt_copper	−0.016	.019	0.08	−0.02	.25	0.46
dt_sodium	−0.026	.0001	0.002	−0.038	.031	0.23
dt_theobromine	.02	.003	0.02	−0.007	.69	0.77
dt_selenium	−0.012	.064	0.21	−0.041	.02	0.20
sup_vite	.008	.22	0.42	−0.043	.014	0.19
sup_thia	−0.009	.17	0.38	−0.041	.02	0.20
sup_ribo	−0.009	.18	0.40	−0.041	.021	0.20
sup_niac	0	.98	0.98	−0.045	.01	0.20
sup_selen	.033	<.0001	<0.0001	.04	.023	0.20
p_veg	−0.015	.023	0.10	−0.04	.025	0.20
p_whitepot	−0.021	.002	0.15	−0.029	.10	0.36
p_veg_oth	−0.019	.006	0.03	−0.034	.05	0.25
p_fruit_oth	.014	.039	0.14	.031	.076	0.30
p_dairy	.02	.003	0.02	.024	.17	0.38
p_milk	.02	.003	0.02	.024	.18	0.38
p_mt_chick_fish	−0.021	.002	0.01	−0.025	.15	0.38
p_beef_pork_lamb	−0.03	<.0001	0.0003	.008	.64	0.75
p_lunch	−0.024	.0004	0.005	−0.036	.044	0.25
p_seafood	.015	.026	0.10	−0.028	.11	0.36
p_soy	.003	.67	0.80	.04	.025	0.20
p_sugar	.022	.0008	0.007	.032	.067	0.28
cal_phos_ratio	.029	<.0001	0.0003	.034	.051	0.25

[a]The identifier annotations of nutrients and food ingredients are the same as those in Table 3.

Prostate cancer prevalence in the AA population is over 1.5 times the prevalence in EA populations. While recent genetic studies have provided explanations for the disparity from multiple angles,[38-41] the etiology of the disparity has been elusive due to the missing heritability problem. Namely, the heritability of PCa susceptibility estimated using AA datasets is much lower than that of the genetic parameter estimated using EA datasets,[1,2] indicating that the contribution of nongenetic factors, including dietary variables and others, to the disease in AA males may be more substantial than that in EA males. In this context, we perceive that beyond those identified here, more PCa- and race-related environmental variables and within-population stratification patterns of diet metrics may exist.

Fourth, we found that of the 80 diet metrics assessed in this study, 28 were associated with blood PSA levels. The positive

correlation between PSA and a few calcium-related metrics, including the intake of dietary calcium and calcium-enriched food ingredients such as milk and the ratio of dietary calcium and phosphorous, was consistent with the results from the National Health and Nutrition Examination Survey data.[42] The moderate negative (or positive) correlation between PSA and protein (or saturated fat) intake confirmed the results in Ref,[43] which showed that men with a lower percentile rank in protein intake and higher percentile rank in fat intake had elevated PSA levels. The positive correlation between PSA and supplementary selenium conflicted with the result from an interventional study, which reported that combined selenium, vitamin E and soy isoflavonoid supplementation reduced serum PSA levels.[44] To our knowledge, the correlations between PSA and the supplementary microminerals copper and zinc have not been reported in the literature. In particular, we noticed that most of the PSA-related diet metrics did not show any association with PCa or were associated with disease risk in the direction opposite to their correlations with PSA levels, indicating that while PSA is a highly useful marker for PCa progression, its associations with diet metrics could hardly be used to identify risk diet/nutrition patterns for PCa. Alternatively, prior to using PSA markers to screen prostate cancer, adjusting for diet metrics may be helpful to increase prediction accuracy.

Acknowledgements

Author Contribution

Study conception: WZ, KZ. Method design: WZ, KZ. Experiments performed: WZ. Data analysis: WZ, KZ. Writing: WZ, KZ. All authors read and approved the final manuscript.

Supplemental Material

Supplemental material for this article is available online.

REFERENCES

1. Mancuso N, Rohland N, Rand KA, et al. The contribution of rare variation to prostate cancer heritability. *Nat Genet*. 2016;48:30-35.
2. Zhang W, Dong Y, Sartor O, Zhang K. Comprehensive analysis of multiple cohort datasets deciphers the utility of germline single-nucleotide polymorphisms in prostate cancer diagnosis. *Cancer Prev Res*. 2021;14:741-752.
3. Nicolosi P, Ledet E, Yang S, et al. Prevalence of germline variants in prostate cancer and implications for current genetic testing guidelines. *JAMA Oncol*. 2019;5:523-528.
4. Kimura T. East meets West: ethnic differences in prostate cancer epidemiology between East Asians and Caucasians. *Chin J Cancer*. 2012;31:421-429.
5. Salmon C, Song L, Muir K, et al. Marital status and prostate cancer incidence: a pooled analysis of 12 case-control studies from the PRACTICAL consortium. *Eur J Epidemiol*. 2021;36:913-925.
6. Rawla P. Epidemiology of Prostate Cancer. *World J Oncol*. 2019;10:63-89.
7. Allen NE, Key TJ, Appleby PN, et al.; Appleby PN. Animal foods, protein, calcium and prostate cancer risk: the European Prospective Investigation into Cancer and Nutrition. *Br J Cancer*. 2008;98:1574-1581.
8. Koh KA, Sesso HD, Paffenbarger Rs Jr, Lee IM. Dairy products, calcium and prostate cancer risk. *Br J Cancer*. 2006;95(11):1582-1585.
9. Rohrmann S, Platz EA, Kavanaugh CJ, Thuita L, Hoffman SC, Helzlsouer KJ. Meat and dairy consumption and subsequent risk of prostate cancer in a US cohort study. *Cancer Causes Control*. 2007;18:41-50.
10. Major JM, Cross AJ, Watters JL, Hollenbeck AR, Graubard BI, Sinha R. Patterns of meat intake and risk of prostate cancer among African-Americans in a large prospective study. *Cancer Causes Control*. 2011;22:1691-1698.
11. Sinha R, Knize MG, Salmon CP, et al. Heterocyclic amine content of pork products cooked by different methods and to varying degrees of doneness. *Food Chem Toxicol*. 1998;36:289-297.
12. Cross AJ, Peters U, Kirsh VA, et al. A prospective study of meat and meat mutagens and prostate cancer risk. *Cancer Res*. 2005;65:11779-11784.
13. Hori SS, Gambhir SS. Mathematical model identifies blood biomarker-based early cancer detection strategies and limitations. *Sci Transl Med*. 2011;3:109ra116.
14. 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;19:558-565.
15. Lewis SJ, Murad A, Chen L, et al. Associations between an obesity related genetic variant (FTO rs9939609) and prostate cancer risk. *PLoS One*. 2010;5: e13485.
16. Noone A, Howlader M, Krapcho N. *SEER Cancer Statistics Review, 1975-2015*. National Cancer Institute; 2018. https://seer.cancer.gov/csr/1975_2015/ (Accessed on August 24, 2021)
17. Gohagan JK, Prorok PC, Hayes RB, Kramer BS, Prostate LC; Ovarian Cancer Screening Trial Project T. The Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial of the National Cancer Institute: history, organization, and status. *Control Clin Trials*. 2000;21(6 Suppl):251S-272S.
18. Shikany JM, Flood AP, Kitahara CM, et al. Dietary carbohydrate, glycemic index, glycemic load, and risk of prostate cancer in the prostate, Lung, colorectal, and Ovarian Cancer Screening Trial (PLCO) cohort. *Cancer Causes Control*. 2011;22:995-1002.
19. Richman EL, Kenfield SA, Stampfer MJ, Giovannucci EL, Chan JM. Egg, red meat, and poultry intake and risk of lethal prostate cancer in the prostate-specific antigen-era: incidence and survival. *Cancer Prev Res*. 2011;4:2110-2121.
20. Aronson WJ, Barnard RJ, Freedland SJ, et al. Growth inhibitory effect of low fat diet on prostate cancer cells: results of a prospective, randomized dietary intervention trial in men with prostate cancer. *J Urol*. 2010;183:345-350.
21. Pelser C, Mondul AM, Hollenbeck AR, Park Y. Dietary fat, fatty acids, and risk of prostate cancer in the NIH-AARP diet and health study. *Cancer Epidemiol Biomarkers Prev*. 2013;22:697-707.
22. Sesso HD, Paffenbarger RS Jr, Lee IM. Alcohol consumption and risk of prostate cancer: the Harvard Alumni Health Study. *Int J Epidemiol*. 2001; 30(4):749-755.
23. Demoury C, Karakiewicz P, Parent ME. Association between lifetime alcohol consumption and prostate cancer risk: A case-control study in Montreal, Canada. *Cancer Epidemiol*. 2016;45:11-17.
24. Kristal AR, Darke AK, Morris JS, et al. Baseline selenium status and effects of selenium and vitamin e supplementation on prostate cancer risk. *J Natl Cancer Inst*. 2014;106(3):djt456.
25. Weinstein SJ, Stolzenberg-Solomon R, Pietinen P, Taylor PR, Virtamo J, Albanes D. Dietary factors of one-carbon metabolism and prostate cancer risk. *Am J Clin Nutr*. 2006;84:929-935.
26. Altman N, Krzywinski M. Points of significance: P values and the search for significance. *Nat Methods*. 2017;14(1):3-4.
27. Tsilidis KK, Allen NE, Appleby PN, et al. Diabetes mellitus and risk of prostate cancer in the European Prospective Investigation into cancer and Nutrition. *Int J Cancer*. 2015;136(2):372-381.
28. Calton BA, Chang SC, Wright ME, et al. History of diabetes mellitus and subsequent prostate cancer risk in the NIH-AARP Diet and Health Study. *Cancer Causes Control*. 2007;18:493-503.
29. 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;176:1121-1129.
30. Watters JL, Park Y, Hollenbeck A, Schatzkin A, Albanes D. Cigarette smoking and prostate cancer in a prospective US cohort study. *Cancer Epidemiol Biomarkers Prev*. 2009;18:2427-2435.
31. Chen X, Zhao Y, Tao Z, Wang K. Coffee consumption and risk of prostate cancer: a systematic review and meta-analysis. *BMJ Open*. 2021;11:e038902.
32. Rider JR, Wilson KM, Sinnott JA, Kelly RS, Mucci LA, Giovannucci EL. Ejaculation frequency and risk of prostate cancer: updated results with an additional decade of follow-up. *Eur Urol*. 2016;70(6):974-982.

33. Barrington WE, Schenk JM, Etzioni R, et al. Difference in association of obesity with Prostate Cancer Risk between US African American and Non-Hispanic white men in the selenium and vitamin E cancer prevention trial (SELECT). *JAMA Oncol*. 2015;1:342-349.

34. Rodriguez C, Freedland SJ, Deka A, et al. Body mass index, weight change, and risk of prostate cancer in the Cancer Prevention Study II nutrition cohort. *Cancer Epidemiol Biomarkers Prev*. 2007;16:63-69.

35. Allott EH, Arab L, Su LJ, et al. Saturated fat intake and prostate cancer aggressiveness: results from the population-based North Carolina-Louisiana Prostate Cancer Project. *Prostate Cancer Prostatic Dis*. 2017;20:48-54.

36. Walker M, Aronson KJ, King W, et al. Dietary patterns and risk of prostate cancer in Ontario, Canada. *Int J Cancer*. 2005;116:592-598.

37. Mocellin S, Briarava M, Pilati P. Vitamin B6 and cancer risk: a field synopsis and meta-analysis. *J Natl Cancer Inst*. 2017;109:1-9.

38. Freedman ML, Haiman CA, Patterson N, et al. Admixture mapping identifies 8q24 as a prostate cancer risk locus in African-American men. *Proc Natl Acad Sci USA*. 2006;103:14068-14073.

39. Robbins CM, Hooker S, Kittles RA, Carpten JD. EphB2 SNPs and sporadic prostate cancer risk in African American men. *PLoS One*. 2011;6:e19494.

40. Paris PL, Kupelian PA, Hall JM, et al. Association between a CYP3A4 genetic variant and clinical presentation in African-American prostate cancer patients. *Cancer Epidemiol Biomarkers Prev*. 1999;8:901-905.

41. Zhang W, Nicholson T, Zhang K. Deciphering the polygenic basis of racial disparities in prostate cancer by an integrative analysis of genomic and Transcriptomic Data. *Cancer Prev Res*. 2022;15:161-171.

42. Song J, Chen C, He S, et al. Is there a non-linear relationship between dietary protein intake and prostate-specific antigen: proof from the national health and nutrition examination survey (2003–2010). *Lipids Health Dis*. 2020;19:82.

43. Ohwaki K, Endo F, Kachi Y, et al. Relationship between dietary factors and prostate-specific antigen in healthy men. *Urol Int*. 2012;89:270-274.

44. Joniau S, Goeman L, Roskams T, Lerut E, Oyen R, Van Poppel H. Effect of nutritional supplement challenge in patients with isolated high-grade prostatic intraepithelial neoplasia. *Urology*. 2007;69:1102-1106.