



RESEARCH

Open Access



HLA diversity unveils susceptibility and organ-specific occurrence of second primary cancers: a prospective cohort study

Zi-Xuan Rong^{1†}, Wei Wei^{1,2†}, Qin Zeng^{1†}, Xiao-Ting Cai^{1†}, Yuan-Yuan Wang³, Jian Wang¹, He-San Luo⁴, Lu-Shan Xiao^{5,6}, Jia-Run Lin¹, Xue Bai¹, Yan-Pei Zhang^{1,5}, Duan-Duan Han¹, Zhong-Yi Dong¹, Wei Wang^{1*}, De-Hua Wu^{1*}  and Si-Cong Ma^{7*} 

Abstract

Background Up to 17% of cancer survivors have been reported to develop second primary cancers (SPC), which cause significant physical and economic distress and often complicate clinical decision-making. However, understanding of SPC remains limited and superficial. Human leukocyte antigen (HLA) is characterized by its polymorphism and has been associated with various diseases. This study aims to explore the role of HLA diversity in SPC incidence.

Methods We analyzed a cohort of 47,550 cancer patients from the UK Biobank. SNP-derived HLA alleles were used and SPC-related HLA alleles were identified using logistic regression, followed by stepwise filtering based on the Akaike information criterion (AIC) and permutation tests. Additionally, we examined the association between extragenetic factors and the risk of SPC in patients carrying hazardous HLA alleles.

Results During a median follow-up of 3.11 years, a total of 2894 (6.09%) participants developed SPC. We identified three protective HLA alleles (DRB1*04:03 and DPA1*02:02 for males and DRB5*01:01 for females) and two hazardous alleles (A*26:01 for males and DPB1*11:01 for females) about SPC. The presence of the protective alleles was associated with a reduced SPC risk (males: hazard ratio [HR] 0.72, 95% confidence interval [CI] 0.59–0.89; females: HR 0.81, 95% CI 0.70–0.93), while the hazardous alleles were linked to an increased risk (males: HR 1.27, 95% CI 1.03–1.56; females: HR 1.35, 95% CI 1.07–1.70). The hazardous allele A*26:01 indicated skin-lung organ-specific SPC occurrence in males. Animal fat and vitamin C were associated with SPC risk in males carrying the hazardous alleles, while free sugar and vegetable fat were linked to SPC risk in females.

Conclusions These results suggest that HLA alleles may serve as biomarkers for the susceptibility and organ-specific occurrence of SPC, while dietary modulation may mitigate hazardous alleles-related SPC risk, potentially aiding in the early prediction and prevention of SPC.

[†]Zi-Xuan Rong, Wei Wei, Qin Zeng and Xiao-Ting Cai contributed equally to this work.

*Correspondence:

Wei Wang

wwei9500@smu.edu.cn

De-Hua Wu

18602062748@163.com

Si-Cong Ma

masc@mail.sysu.edu.cn; masicong97@foxmail.com

Full list of author information is available at the end of the article



Keywords Second primary cancer (SPC), Human leukocyte antigen (HLA), Cancer immunity, Cancer prevention, Dietary intervention

Background

Due to great achievements in screening, diagnosis, and treatments, the survival of cancer patients has greatly improved. However, this encouraging trend has been faced by a new challenge, namely second primary cancers (SPC), which is known as a late effect of cancer [1]. SPC is generally defined as the occurrence of more than one cancer in the same individual, which has been reported in up to 17% of cancer survivors [2]. There is heterogeneity in the timing of SPC occurrence; the Surveillance Epidemiology and End Results (SEER) recommended discriminating metachronous SPC (the second cancer was diagnosed more than 2 months after the first cancer diagnosis) from synchronous one (the second cancer was diagnosed within 2 months of the first cancer diagnosis). SPC usually causes great physical and economic distress and often complicates clinical decision-making. Moreover, patients with an SPC cancer history are generally excluded from most clinical trials, robbing their privilege for optimal treatments [2]. To date, our understanding of SPC has largely been limited to broad demographic analyses. While these analyses are essential for comprehending SPC, there remain no clear methods or convenient tools for predicting or preventing the occurrence of SPC. Given the high incidence rate and significant risk of SPC, a deeper understanding is essential. There is a need for a comprehensive system that integrates accurate prediction and intervention strategies to reduce the risk of SPC.

Human leukocyte antigen (HLA) is located on the short arm of human chromosome 6, representing the most sophisticated genetic system in the human genome. Decades of research on HLA have confirmed its dominant role in human immunity, both protective and deleterious [3]. HLA is characterized by its polymorphism, manifesting as different HLA alleles. The HLA region on the short arm of chromosome 6 was divided into several smaller regions, including the eleven regions in UK Biobank dataset (i.e., A, B, C, DRB1, DRB3, DRB4, DRB5, DQA1, DQB1, DPA1, DPB1) [3]. Interestingly, HLA alleles have been considered as key factors in understanding a wide range of diseases, including ankylosing spondylitis, multiple sclerosis, and COVID-19, all of which share an immunological component [3–7]. Immunity also plays a critical role in tumorigenesis and tumor development. A recent multi-cohort study also demonstrated that the HLA allele HLA-A*03 could predict poor response to immune checkpoint blockade across several cancers, further supporting the potential role of

different HLA alleles in cancer immunity [8]. Given the close relationship between HLA and cancer, it is logical to consider that HLA alleles may serve as a key factor in understanding SPC. The independent effect of a single allele on immune and cancer-associated outcomes has been indicated in several studies [5, 9–11]. To avoid complexity and enhance future applicability, we focused on the association between SPC risk and individual HLA alleles rather than their combinations, aiming at linking genetics and immunity. Beyond inherited factors, growing evidence suggested that extragenetic risk factors, like metabolic and behavioral influences, could worsen the risk of cancer incidence and mortality [12]. However, the extent to which these extragenetic factors may offset the effects of hazardous HLA alleles remains largely underestimated. Targeting these extragenetic factors may mitigate the inherent limitations of genetic factors.

Based on this premise, we hypothesized that combining risk prediction through HLA alleles and interventions on extragenetic factors could together reduce the risk of SPC. To this end, we distinguished between protective and hazardous HLA alleles in a large prospective cohort in the UK, examined their ability to indicate organ-specific SPC occurrence, and identified extragenetic factors that might mitigate the detrimental effects of hazardous HLA alleles.

Methods

Cohort resources

The UK Biobank program is a large population-based prospective study, aiming to investigate the genetic and nongenetic determinants of health-related outcomes of people aged 40–69 years [13–15]. Recruitment of more than 500,000 participants was completed during 2006–2010. Owing to the fine design and big effort, UK Biobank successfully surmounted data depth and data breadth, two common shortcomings of prospective studies, and therefore guaranteed a high quality of the data.

Participants of the study

The UK Biobank project offered health outcomes tracked by linking to disease registry databases from the UK National Health Service, including cancer registry [16]. Participants with inconsistent self-reported sex from genetic sex ($n=372$), which may represent low-quality data, withdrawing consent ($n=158$), and with cancer history before attendance ($n=23,803$), which may confound baseline characteristics, were excluded, resulting

in 50,164 participants with cancer records after attendance. Within cancer registry, “Type of cancer: ICD10” (Field ID: 40006) and “Behaviour of cancer tumor” (Field ID: 40012) were used to identify non-malignant tumors and metastatic tumors from primary malignant tumors. Malignant tumor was coded with “C,” while benign, in situ, or uncertain behavior tumors were coded with “D” in ICD10. “Uncertain whether benign or malignant,” “Malignant, metastatic site,” and “Malignant, uncertain whether primary or metastatic” were excluded in “Behavior of cancer tumor”. Non-malignant cancer, metastatic cancer, incomplete cancer or HLA information were excluded ($n=2614$), finally resulting in 47,550 participants for further analyses. Our analysis focused on 47,550 cancer patients who were initially diagnosed with cancer only after attendance.

Second primary cancer ascertainment in UK Biobank

The primary outcome of the study was the incidence of subsequent primary malignant tumors. Data was collected from national cancer registries. Cancer-related data included cancer type, date, metastasis, and behavior. Cancer type was coded using The International Statistical Classification of Diseases and Related Health Problems 10th Revision (ICD10), within which malignant neoplasms were coded from C00 to C97. In our study, we excluded non-melanoma skin cancer (C44) and not-well-defined primary or metastatic tumors (C76–C80) and included all other well-defined malignant primary tumors. SPC was defined as more than one cancer in the same individual. Precisely, in our study, we used ICD10 top-level code to identify a cancer. A second cancer with non-metastatic indication and malignant nature and with different ICD codes from the first one was defined as SPC. The ending timepoint was Feb 29, 2020, for England and Wales and Jan 31, 2021, for Scotland.

HLA allele status ascertainment

The four-digit HLA data were from SNPs imputation data using the HLA*IMP:02 algorithm and available as a table of HLA*IMP:02 imputations in the “Genome” section of UK Biobank. Thus, HLA alleles represent genetic-level evidence. The value referred to the absolute posterior probability of the allele inference. We adopted the notes for researchers using 0.7 as a threshold to discriminate the presence from the absence of an HLA allele (<0.7 for absence while ≥ 0.7 for presence). Additionally, to avert fortuity of the conclusion, only HLA alleles with frequency $>0.1\%$ were included in the analysis. To explore the combinations, we adopted an empirical and relatively liberal threshold of 1 and no other present alleles in the region to indicate the same two alleles in an individual. Since the accuracy and call rates of the imputation

algorithm were not 100% and the inference was based on a setting threshold, there would be missing alleles in several participants. Only individuals with complete imputation in all eleven regions were included in the combination analysis. Notably, for HLA-DRB3, -DRB4, and -DRB5 the allele “99:01” indicates that no allele is present, because these loci can have copy number 0. They were not included in the allele frequency description. A total of 364,677 participants had complete imputation at all eleven sites.

Screening of protective versus hazardous HLA alleles concerning SPC

We adopted the strategy below to screen the SPC-related protective and hazardous HLA alleles: (a) univariate logistic regression model was used to narrow the size of the candidate alleles with a relatively loose P value <0.1 ; (b) both-direction stepwise process to optimize the model based on minimal Akaike information criterion (12,752 for males and 8928 for females), which indicates the current HLA alleles make the least “loss of message”; (c) significant HLA alleles were then sent to multivariable logistic regression. Alleles with a P value less than 0.05 were selected. Additionally, we conducted a permutation test to evaluate whether the observed statistic is significantly different from the expected value under random conditions by random permutation of samples. The permutation test was implemented using “glmpermu” R package (version 0.0.1) with the seed set to “1234” for replicability, setting permutation times to 1000 as default. Final adjusted permutation P value after 1000 permutation times for included hazardous and protective alleles were reported (Additional file 1: Table S1). Alleles with a permutation P value less than 0.05 were considered SPC-related hazardous or protective alleles, depending on the beta value from the multivariable logistic regression model.

The association between hazardous and protective alleles and SPC

Accumulative risks of SPC divided by per HLA allele status and gender were shown by cumulative risk curves and the log-rank test was applied for statistical analysis. The HLA status was set to binary variates that indicated the presence or not of hazardous or protective alleles. To adjust the heterogeneity of the real-world cohort with such a large sample size, we applied additional multivariable Cox regression models: (1) model 1 was adjusted for age and ethnicity; (2) model 2 was additionally adjusted for UK Biobank assessment center; (3) model 3 was additionally adjusted for Townsend Deprivation Index (TDI) which reflected the socioeconomic status. The hazard ratio (HR), 95% confidence interval (95% CI), and P value

were reported. R package “survival” (version 3.5–7) and “survminer” (version 0.4.9) were used.

Organ-specific association between HLA alleles and SPC

Fisher’s exact test was used to investigate the contingency between hazardous alleles and different first cancers in the context of several selected SPCs with a higher incidence than random background. The Benjamini and Hochberg adjustment was applied for multiple comparisons. Univariable logistic regression model was used to validate the relationship between A*26:01 and second lung cancer incidence with first skin cancer. Functions “fisher.test()” and “p.adjust()” from R package “stats” (version 4.2.2) were used.

Potential interventions of extragenetic factors

The association between behavioral (smoking and alcohol drinking status, and physical activity), sociopsychological (depression, loneliness, and social isolation), and alimentary fields (sugar, fat, protein, and vitamin intake) and SPC were analyzed. Smoking and alcohol drinking status was classified into “current” and “non-current” (never or previous). Depression was assessed using a two-item Patient Health Questionnaire (PHQ-2) at attendance. Participants were asked two questions about the frequency of depressive feelings and anhedonia in the past 2 weeks. Answers of “not at all,” “several days,” “more than half the days,” and “nearly every day” scored 0, 1, 2, and 3 points, respectively. A PHQ-2 score ≥ 3 points is considered depression. Loneliness was assessed by inquiring about the feelings of loneliness (“yes” for 1 point while “no” for 0) and the frequency of confiding (more than once a month for 0 point while others for 1). A sum score of 2 indicated loneliness. Social isolation was assessed by inquiring about the household members (living alone for 1 point), frequency of visiting with friends or family (less than once a month for one point), and leisure activities attending (none for one point). Participants with a total score of more than two points were considered in social isolation status. Moderate physical activity > 675 metabolic equivalents (MET) minutes per week and vigorous > 450 MET minutes per week was considered preferable. An online follow-up program was held by UK Biobank, sending questionnaires to collect the dynamic diet changes. Nutrient intake was estimated based on 24-h recall of the previous day. The average estimated intake was used to reflect the prolonged nutrient intake. To characterize the best intervention cut-point of potential nutrient intake, the best cutoff value calculated by the “survminer” R package was used for dichotomization to low and high groups (Additional file 2: Table S1). Cox proportional hazard model was used to test the association of extragenetic factors with SPC. We first attempted

the univariable Cox model, followed by multivariable Cox model including all variables with $P < 0.1$ in the univariable one. The hazard ratio (HR), 95% confidence interval (95% CI), and P value were reported.

Results

Study population from UK Biobank

A total of 47,550 participants of the UK Biobank program were included in analyses (Fig. 1). The sociodemographic characteristics of enrolled participants were shown in Additional file 3: Table S1. During a median follow-up of 3.11 years, 2894 (6.09%) participants developed second primary cancers, among whom 1763 (60.92%) were males and 1131 (39.08%) were females. Males possessed a higher risk of SPC (HR 1.53, 95% CI 1.42–1.64; $P < 0.0001$), compared with females (Additional file 4: Fig. S1). Since SEER recommended discriminating metachronous SPC (≥ 2 months) from synchronous one, a total of 2296 (79.34%) cases were classified into metachronous group in the present cohort. The frequency of HLA alleles in the population was shown in Additional file 5: Table S1. The most common HLA allele in the total population was DPA1*01:03. Interestingly, it was also more prevalent in individuals with cancers compared to those without cancers (adjusted $P = 0.01$ in males and $P = 0.0006$ in females) (Additional file 5: Table S1 and Additional file 6: Tables S1–11). However, it could not differentiate SPC patients from non-SPC patients. The combinations of HLA alleles were divergent as expected, with totally 263,925 different HLA combinations under 4-digit resolution (Additional file 7: Table S1).

Screening of SPC-related protective and hazardous HLA alleles

To investigate the relationship between HLA alleles and SPC, univariate logistic regression followed by bidirectional stepwise filtering was performed, yielding the selection of several alleles for multivariable model. Permutation tests were subsequently applied to these alleles (Fig. 2A). The positive beta value in the model defined hazardous HLA alleles while the negative defined protective ones. Two protective alleles (DRB1*04:03, permutation $P = 0.001$ and DPA1*02:02, permutation $P = 0.002$) and one hazardous allele (A*26:01, permutation $P = 0.031$) of SPC were identified for males, while one protective allele (DRB5*01:01, permutation $P < 0.0001$) and one hazardous allele (DPB1*11:01, permutation $P = 0.036$) for females (Additional file 1: Table S1). The frequency of the male hazardous allele A*26:01 was 43.06% in cancer patients and 41.97% in those without cancer (adjusted P value = 0.93). Similarly, for the female hazardous allele DPB1*11:01, no significant difference was found between the cancer and cancer-free groups

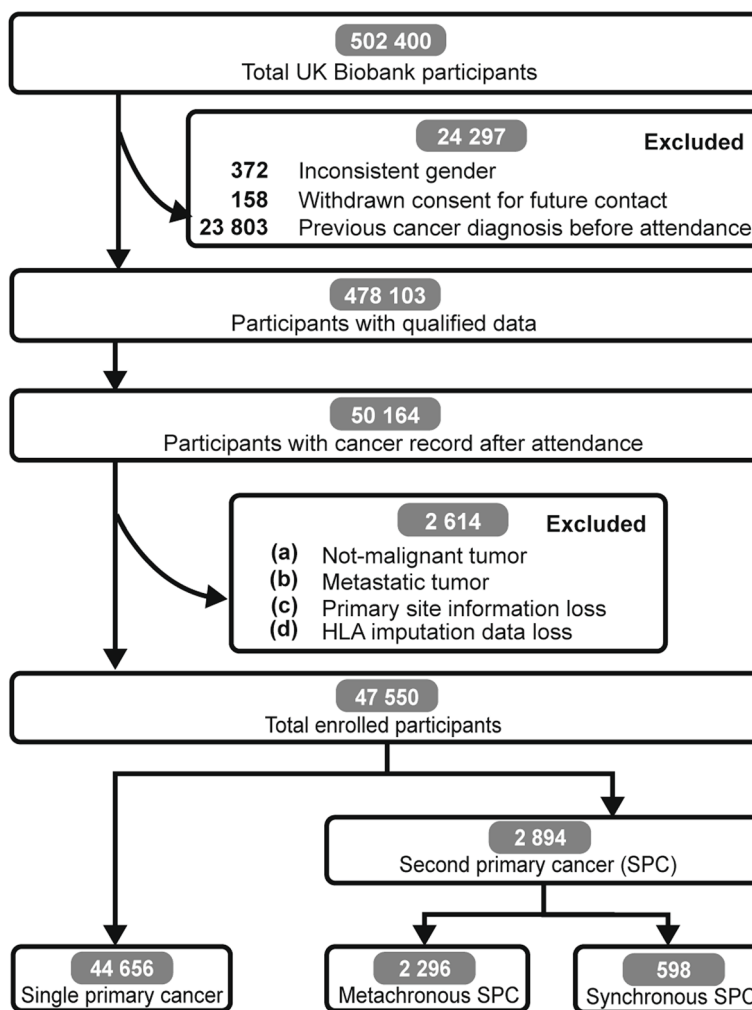


Fig. 1 Flowchart showing the enrollment process of participants. 478,103 participants were included after basic quality control process. 47,550 cancer patients were finally enrolled for analyses based on certain conditions

(adjusted P value=0.41). Similar results were observed for protective alleles, although DPA1*02:02 was more prevalent among cancer-free male participants (Additional file 6: Tables S1, S4, S5, S8, and S11). Overall, there were no significant differences between the cancer and cancer-free groups for most hazardous and protective alleles, suggesting that these alleles may be specific to SPC.

The capacity of the identified HLA alleles for SPC prediction was subsequently validated. The presence of protective alleles was associated with a lower risk of SPC compared to individuals without protective alleles for males (HR 0.72, 95% CI 0.59–0.89; $P=0.0019$) and females (HR 0.81, 95% CI 0.70–0.93; $P=0.0029$). Conversely, the presence of hazardous alleles was linked to an increased risk of SPC for males (HR 1.27, 95% CI 1.03–1.56; $P=0.023$) and females (HR 1.35, 95% CI 1.07–1.70; $P=0.011$), compared to those without these

alleles (Fig. 2B and C). After adjusting several sociodemographic covariates, the estimates of hazardous and protective alleles on SPC were still plausible (Additional file 8: Fig. S1). Collectively, these findings demonstrate that HLA alleles are associated with the risk of SPC and can effectively stratify cancer patients into distinct SPC risk groups.

HLA alleles indicate organ-specific occurrence of metachronous SPC

Metachronous SPC has a delayed onset which provides an opportunity for prediction and prevention. Therefore, we focused on patients whose second cancer developed more than 2 months after the first cancer diagnosis, aiming to identify the factors associated with metachronous incidence. Eleven and thirteen common cancers classified by affected organs were selected for organ-specific analysis for males and

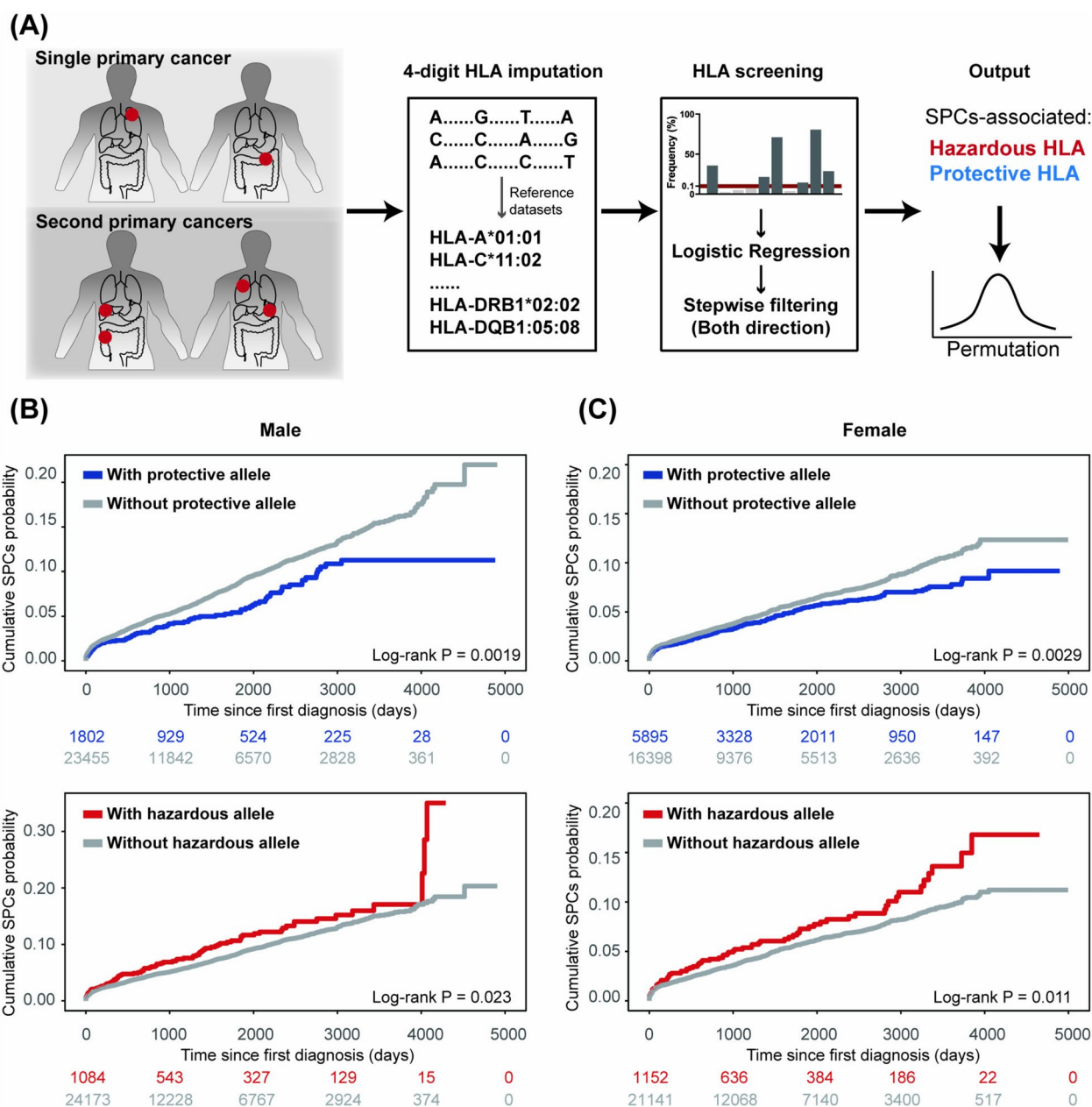


Fig. 2 Screening of SPC-related protective and hazardous HLA alleles. **A** Schematic illustrating the screening process. More details could be found in the method section. Cumulative SPC incidence of **B** males and **C** females stratified by HLA alleles. Log-rank test was used for statistical analysis. Hazardous alleles: A*26:01 for males and DPB1*11:01 for females. Protective alleles: DRB1*04:03 and DPA1*02:02 for males and DRB5*01:01 for females

females, respectively (Fig. 3A and B and Additional file 9: Table S1). The SPC profile showed that prostate, breast, lung, hematopoietic, and colorectal cancers were the most susceptible second cancers in males and females. Although prostate and breast cancers were the most common first cancers in men and women, respectively, their proportions significantly declined in second cancer diagnoses. In contrast, the incidence of second

lung, urinary, hematopoietic, and pancreatic cancers increased notably compared to the initial diagnoses, regardless of gender (Fig. 3C and D). Taken together, these results suggested that metachronous SPC was not merely a result of random incidence. Rather, there may be some underlying factors driving the metachronous association that warrant further investigation.

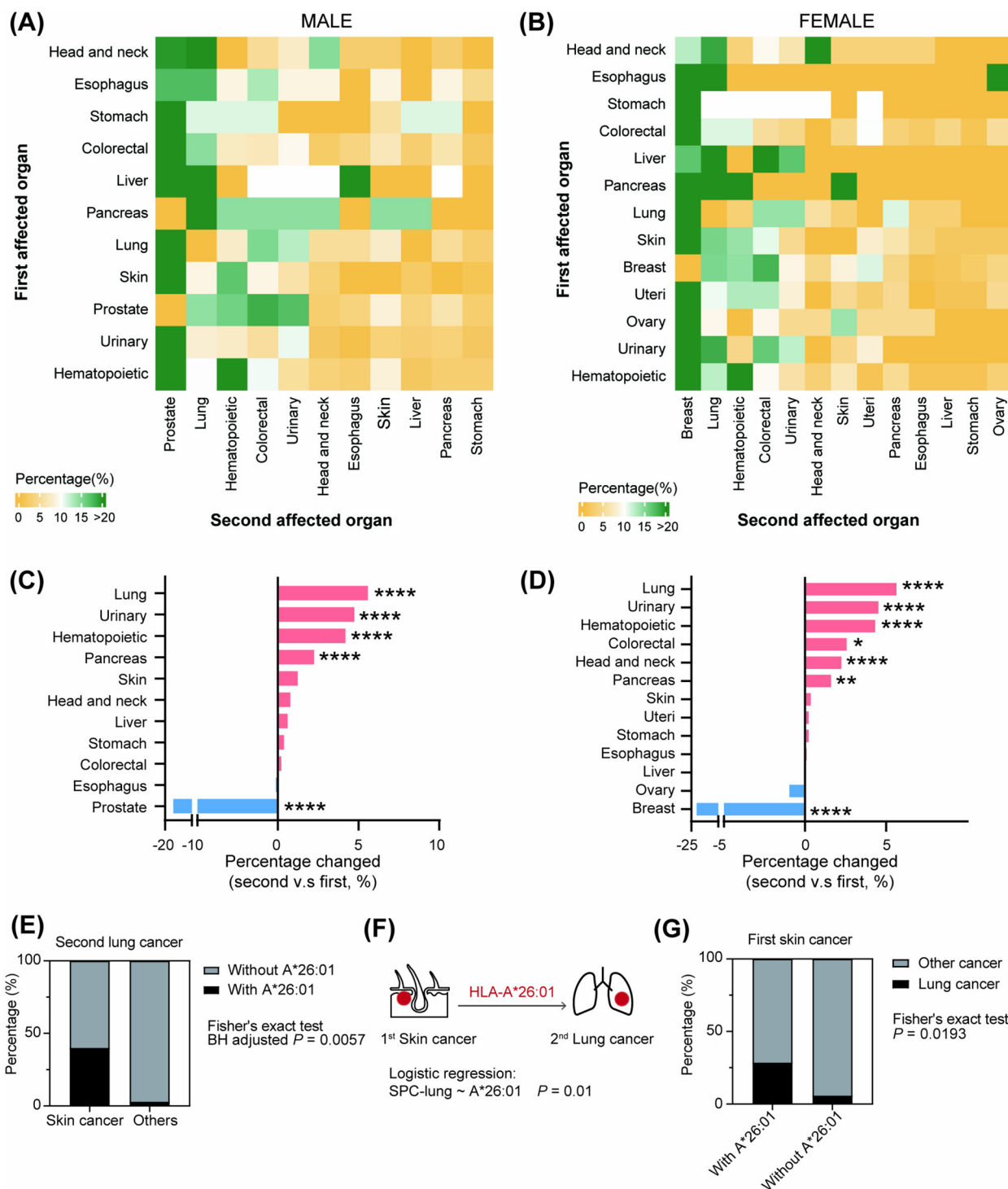


Fig. 3 Profile of the first and second cancers and organ-specific HLA indication. **A–B** Heatmap showing the distribution of subsequent cancers. **C–D** Bar plot showing the changes of constitute ratio in the subsequent setting versus the first setting. The value was presented as the percentage in the second setting minus that in the first. Chi-square test was used for statistical analysis. * $P < 0.05$, ** $P < 0.01$, and **** $P < 0.0001$. **E** Frequency of A*26:01 in second lung cancer male patients divided by first cancer (skin versus others). The Fisher's exact test was applied with BH multiple comparisons. **F** Schematic model indicating the role of A*26:01 in skin-to-lung organ-specific SPC incidence. Logistic regression was applied. **G** Frequency of second lung cancer in first skin cancer patients divided by A*26:01 status

Given that hazardous alleles could increase the risk of SPC and that certain SPCs exhibited higher incidence than expected by chance, we then investigated whether there were some associations between hazardous alleles and the first cancers that contributed to specific second cancers. To achieve this goal, we focused on cancers with second incidence rates exceeding the random background. Results showed that among male patients with second lung cancers, those with first skin cancer had a higher proportion of allele A*26:01 (adjusted $P=0.0057$) (Fig. 3E). Furthermore, patients with first skin cancer showed a significantly stronger association with second lung cancer when carrying the A*26:01 allele (Fig. 3F). Additionally, the frequency of patients with first skin cancer developing second lung cancer was significantly higher among those carrying allele A*26:01 than those without it ($P=0.0193$, Fig. 3G). These results implied that HLA allele A*26:01 could, at least partly, explain the skin-to-lung organ-specific metachronous association.

A holistic system to minimize SPC hazard through genetic prediction and extragenetic intervention

Based on the above results, we could predict the SPC-susceptible population using HLA alleles. Carriers of hazardous HLA alleles are at a higher risk of developing SPC. In this seminar, we recommended that patients follow specific guideline to reduce their SPC risk, which is more frequent and precise medical examinations if hazardous HLA alleles are detected, rather than relying solely on routine screenings. Additionally, various extragenetic factors have been implicated in cancer development. To investigate the factors influencing cancer patients' risk of subsequent cancers, we analyzed data related to socio-psychological, behavioral, and dietary aspects (Fig. 4A).

For male patients carrying hazardous HLA allele, higher animal fat intake was associated with an increased SPC risk (HR, 2.44, 95% CI 1.23–4.84; $P=0.01$) while higher vitamin C intake correlated with a reduced SPC risk (HR, 0.41, 95% CI 0.2–0.84; $P=0.01$) after multivariable adjustment. In female patients, sensitivity to free sugar and vegetable fat was noted. Higher free sugar intake was linked to an increased SPC risk (HR, 2.32, 95% CI 1.04–5.18; $P=0.04$) whereas higher vegetable fat intake was associated with a decreased SPC risk (HR, 0.35, 95% CI 0.15–0.80; $P=0.01$) (Fig. 4B). These results reveal that there is heterogeneity in dietary susceptibility between male and female cancer patients carrying hazardous SPC-related HLA alleles. We also observed that male patients with hazardous alleles have lower level of lymphocyte count ($P=0.02$), but higher level of peripheral neutrophil count ($P=0.05$), monocyte count ($P=0.003$), and neutrophil–lymphocyte ratio (NLR) ($P=0.002$), relative to those with protective alleles,

whereas female patients with hazardous alleles have higher level of neutrophils count ($P=0.02$) and eosinophils count ($P=0.003$), relative to those with protective alleles (Additional file 10: Fig. S1).

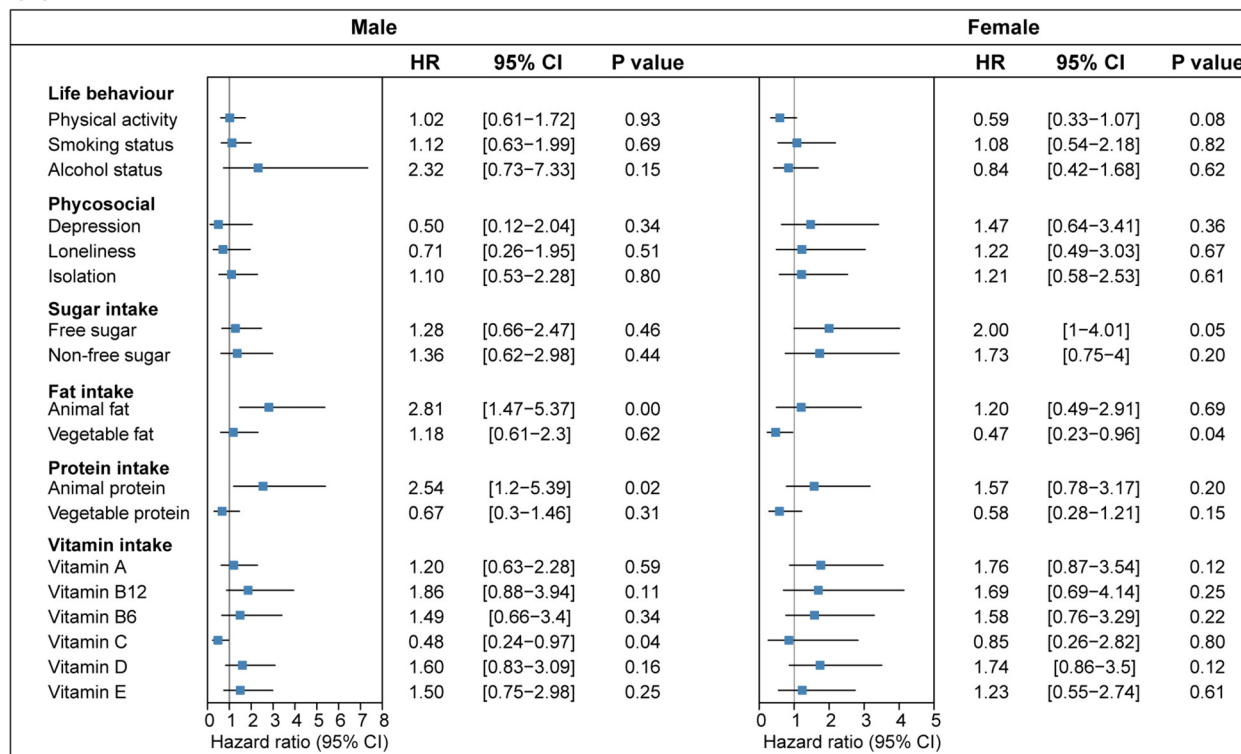
Along with HLA prediction, parallel modifications to dietary patterns might contribute to a lower risk of SPC. A Mediterranean-like or Dietary Approaches to Stop Hypertension (DASH)-like pattern is recommended, emphasizing the reduction of animal fat intake and encouraging the use of vegetable oils. Furthermore, for females, minimizing free sugar intake is advisable. Taken together, we proposed an HLA-diet-oriented system that integrated genetic prediction with extragenetic intervention to mitigate the risk of SPC, which might help to achieve the epidemiological goal of primary and secondary prevention of SPC (Fig. 5).

Discussion

In this large population-based prospective cohort study, we explored the relationship between SPC and both genetic HLA factors and extragenetic factors. We found that HLA alleles could predict the risk of SPC incidence. Further investigation into the intrinsic associations of metachronous SPC revealed that the hazardous HLA allele A*26:01 for males was associated with the organ-specific incidence of first-skin and second-lung cancers. Analysis of extragenetic factors identified potential dietary interventions for those carrying hazardous HLA alleles. Together, these findings support the development of a conceptual HLA-diet-oriented model for the early detection and intervention of SPC.

We described the association between SPC risk and HLA alleles. Although SPC has been a heated but old topic, dating back to the year 1921, most previous studies were limited to the epidemiological field, lacking a deeper insight [1, 17–22]. HLA is a gene-level factor and different HLA alleles have been confirmed as susceptible factors of several diseases [3–7]. Classical HLA class I and II molecules, coded by HLA-A, -B, -C and -DP, -DQ, -DR genes, mainly function vitally in the antigen peptides presentation to CD8+ T cells and CD4+ T cells, thereby yielding an efficient response to fight against infection and malignancy while maintaining self-tolerance. A recent multi-cohort study regarding cancer immune checkpoint blockade mainly affecting T cells revealed that HLA-A*03 predicted poor response to ICB [8]. HLA shares an immunological component with oncogenesis and therefore may lead to a potential opportunity to decipher SPC. Several studies demonstrated gender bias in the association between HLA alleles and diseases [23–26]. Likewise, the gender difference in physiology and cancer spectrum requires a separate discussion on the association between HLA and SPC. Indeed, we screened out

(A)



(B)

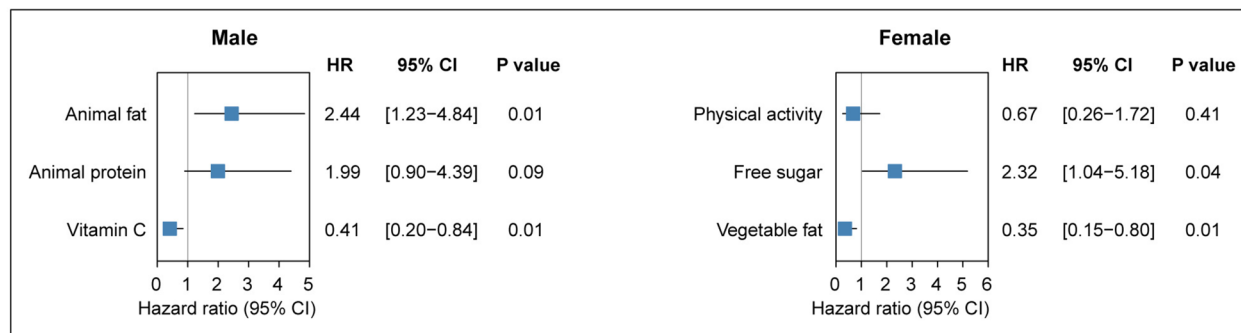


Fig. 4 Forest plot indicating the association between extragenetic factors and SPC. The association between extragenetic factors and SPC incidence estimated by **A** univariable and **B** multivariable Cox proportional hazard models. Reference group: physical activity (preferable); smoking and alcohol status (non-current); psychosocial (absence); sugar, fat, protein, and vitamin intake (low intake)

two protective and one hazardous HLA alleles for males, while one protective and one hazardous HLA alleles for females, whose capacity to predict SPC risk was further validated. It is believed that HLA-peptide-TCR interactions are regulated by a set of perturbed rules, such as atypical docking, low-affinity peptide binding, stabilization of weak peptide-HLA, and altered register [3]. These mechanisms may provide clues to give insights into SPC. Given that HLA is highly correlated with immunity, we also explore the association between peripheral immune

cells and HLA alleles. Previous studies have indicated the unfavorable role of NLR on cancer incidence risk [27] and prognosis [28]. Higher monocyte count was reported associated with poor prognosis in several cancers [29–32]. However, a favorable role of higher eosinophil counts on cancer immunity has been indicated in previous studies [33, 34]. Whether these differences can explain SPC risk through hazardous and protective alleles remains to be validated in larger prospective cohort studies and

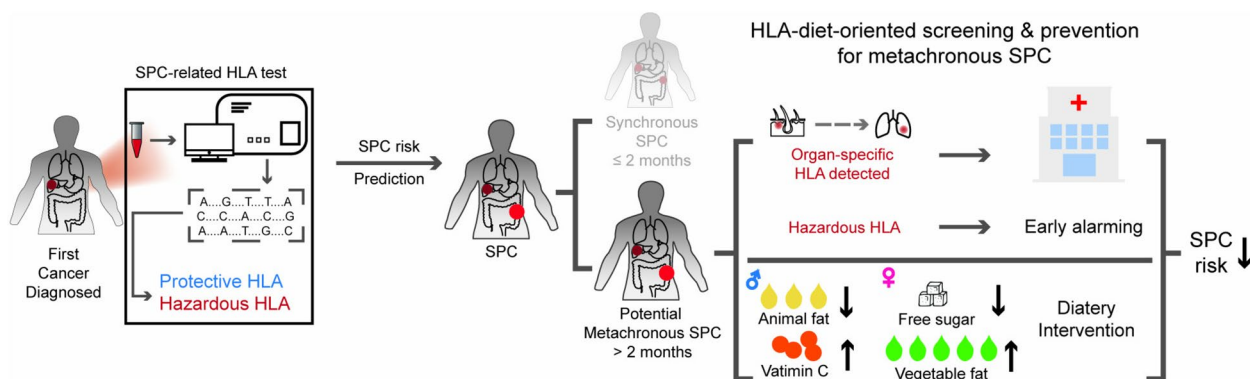


Fig. 5 Schematic model illustrating the HLA-diet-oriented system with genetic prediction and extragenetic intervention against SPC

mechanistic research. Therefore, these results should be interpreted with caution.

Compared with synchronous SPC, metachronous SPC is possible for clinical intervention. In this study, there were 223 male and 214 female SPC patients with more than two primary cancers. To investigate the intrinsic mystery and potential intervention strategies, we focused on metachronous SPC. International Association of Cancer Registries and International Agency for Research on Cancer (IACR/IARC) recommended a 6-month cut-point to classify metachronous and synchronous SPC yet SEER recommended 2 months. We adopted the latter to magnify the representativeness of the population. In fact, the Siriraj Cancer Center cohort study showed that most of second cancers (70.87%) were diagnosed more than 2 months after the initial cancer diagnosis, which aligned closely with our data (79.34%) [22]. Differences in the constituent ratio of cancer in the first and second settings implied that there may be some intrinsic association between the first and the second affected systems. In line with a prior study of a Thailand cohort, we also found a high metachronous association of prostate cancer with second colorectal cancers [22]. Remarkably, one of our previously identified hazardous HLA alleles was found to indicate organ-specific SPC incidence, again inferring the denoting value of HLA polymorphism in diseases. This may be attributed to the immunopathology caused by opportune exposure of specific susceptible HLA alleles to specific microenvironments under multiple conditions.

To pursue convenient interventions for those with hazardous HLA alleles, we performed analyses between common extragenetic factors and SPC. Different dietary patterns were shown to be associated with SPC risk. Higher animal fat intake was associated with increased SPC risk in males, while higher vegetable fat intake was associated with decreased SPC risk in females. Besides, higher vitamin C intake reduced SPC risk in males. Such

low-animal-fat and qualified-vegetable-fat diet resembles Mediterranean Diet and DASH Diet. Mediterranean Diet advocates a high intake of food rich in vegetable fat but a low intake of red meat, producing fewer oxidative and inflammatory factors while positively influencing gut microbiota to support the immune system [35, 36]. Female patients also seemed to be more sensitive to free sugar. Sugar could be classified into free sugar and non-free sugar; free sugar refers to all monosaccharides and disaccharides added to food as well as those naturally present in honey, syrups, and fruit juices [37]. Although WHO recommended less than 10% of the daily energy intake of free sugar and some studies did show lower free sugar intake correlated with a lower risk of certain cancers, there was still some antagonistic voice [38]. However, in our study, female patients carrying hazardous HLA alleles were shown to benefit from lower free sugar intake against SPC, which, although, needs further validation.

It is noteworthy that our study has some strengths. First, the 47,550 cancer patients' data was derived from a well-designed and well-managed prospective cohort, covering extensive information at all levels. Second, our results indicated a new approach to understanding SPC, which represented the essential genetic level evidence. Finally, we proposed a novel model that integrated genetic and extragenetic factors to achieve the maximum reduction in SPC risk.

Nevertheless, we have to acknowledge several limitations of our study. First, most of the ethnicity of participants is white, limiting the generalization of the results. Second, some possible factors that may affect SPC incidence like different histological types, gene mutations, and treatments were not considered. Third, selection bias and recall bias are common in such programs, which may impair the authenticity of the result. Fourth, most of the extragenetic data were collected at attendance,

which may not reflect the situation after the first diagnosis. Fifth, cancer diagnosis through a code-based way is of dampened precision. Sixth, allele status was produced based on the algorithm, so the extension of the conclusion using other ways to test HLA requires attention. Finally, the study was designed as a single cohort, so the conclusions need to be interpreted carefully and further cohort studies are necessary to validate the conclusions.

Conclusions

In conclusion, our study indicates that certain HLA alleles are associated with SPC risk and some even predict organ-specific occurrence. Lower animal fat intake and higher vitamin C intake for males, as well as reduced free sugar intake and sufficient vegetable fat intake for females carrying hazardous HLA alleles, may help reduce their risk of second cancers. The HLA-diet-oriented prediction and prevention model might help mitigate the public cancer challenge.

Abbreviations

AIC	Akaike information criterion
CI	Confidence interval
DASH	Dietary Approaches to Stop Hypertension
HLA	Human leukocyte antigen
HR	Hazard ratio
ICD10	International Classification of Diseases version 10
IACR/IARC	International Association of Cancer Registries and International Agency for Research on Cancer
MET	Metabolic equivalents
PHQ-2	Two-item Patient Health Questionnaire
SPC	Second primary cancers
SEER	The Surveillance Epidemiology and End Results

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12916-024-03676-6>.

Supplementary Material 1.
Supplementary Material 2.
Supplementary Material 3.
Supplementary Material 4.
Supplementary Material 5.
Supplementary Material 6.
Supplementary Material 7.
Supplementary Material 8.
Supplementary Material 9.
Supplementary Material 10.

Acknowledgements

This research has been conducted using the UK Biobank resource under application number 92668. We are grateful to UK Biobank participants, as well as all the research staff who worked on the data collection and management for their contribution.

Authors' Contribution

Conceptualization: Z.X. Rong, Z.Y. Dong & S.C. Ma; Methodology: Z.X. Rong, Z.Y. Dong & S.C. Ma; Data curation: Z.X. Rong, J.R. Lin, Z.Y. Dong & S.C. Ma; Formal

analysis: Z.X. Rong, W. Wei, Q. Zeng & X.T. Cai; Funding acquisition: D.H. Wu, Z.Y. Dong, X. Bai & W. Wang; Investigation: Y.Y. Wang, J. Wang, H.S. Luo, L.S. Xiao, X. Bai, Y.P. Zhang & D.D. Han; Visualization: Z.X. Rong, W. Wei, Q. Zeng & X.T. Cai; Writing-original draft: Z.X. Rong, W. Wei, Q. Zeng & X.T. Cai; Writing-review & editing: Z.X. Rong, Z.Y. Dong, W. Wang, D.H. Wu & S.C. Ma. All authors read and approved the final manuscript.

Funding

This study was supported by the Regional Joint Fund Project of Basic and Applied Basic Research Foundation of Guangdong Province (Grant No. 2022B1515120035), the National Natural Science Foundation of China (Grant No. 82272820 and 82272731), the Outstanding Youths Development Scheme of Nanfang Hospital, Southern Medical University (Grant No. 2023J002), President Foundation of Nanfang Hospital, Southern Medical University (Grant No. 2021A004), and College Students' Innovative Entrepreneurial Training Plan Program (Grant No. 202212121019 and 202212121016). The funder has no roles in the conceptualization, design, data collection, analysis, decision to publish, or preparation of the manuscript.

Availability of data and materials

The access to UK Biobank can be gained by registration followed by application at <https://www.ukbiobank.ac.uk/>. This research has been conducted using the UK Biobank resource under application number 92668.

Declarations

Ethics approval and consent to participate

The UK Biobank received ethical approval from the Research Ethics Committee (REC reference for UK Biobank 11/NW/0382) and all the enrolled participants provided written informed consent. Institutional review board approval was waived because the data is anonymous and publicly available.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of Radiation Oncology, Nanfang Hospital, Southern Medical University, 1838 North Guangzhou Avenue, Guangzhou 510515, Guangdong, China. ²Department of Oncology, Xiangyang Central Hospital, Hubei University of Arts and Science, Xiangyang 441000, Hubei, China. ³Department of Oncology, Nanfang Hospital, Southern Medical University, Guangzhou 510515, Guangdong, China. ⁴Department of Radiation Oncology, Shantou Central Hospital, Shantou 515031, Guangdong, China. ⁵Information Management and Big Data Center, Nanfang Hospital, Southern Medical University, Guangzhou 510515, Guangdong, China. ⁶Guangdong Provincial Key Laboratory of Viral Hepatitis Research, Hepatology Unit and Department of Infectious Diseases, Nanfang Hospital, Southern Medical University, Guangzhou 510515, Guangdong, China. ⁷Department of Thoracic Surgery, The First Affiliated Hospital, Sun Yat-Sen University, 58 Zhongshan 2nd Road, Guangzhou 510080, Guangdong, China.

Received: 8 March 2024 Accepted: 1 October 2024

Published online: 08 October 2024

References

- Keegan THM, Bleyer A, Rosenberg AS, Li Q, Goldfarb M. Second primary malignant neoplasms and survival in adolescent and young adult cancer survivors. *JAMA Oncol.* 2017;3(11):1554–7.
- Vogt A, Schmid S, Heinimann K, Frick H, Herrmann C, Cerny T, et al. Multiple primary tumours: challenges and approaches, a review. *ESMO Open.* 2017;2(2): e000172.
- Dendrou CA, Petersen J, Rossjohn J, Fugger L. HLA variation and disease. *Nat Rev Immunol.* 2018;18(5):325–39.

4. Langton DJ, Bourke SC, Lie BA, Reiff G, Natu S, Darlay R, et al. The influence of HLA genotype on the severity of COVID-19 infection. *HLA*. 2021;98(1):14–22.
5. Karnes JH, Bastarache L, Shaffer CM, Gaudieri S, Xu Y, Glazer AM, et al. Phenome-wide scanning identifies multiple diseases and disease severity phenotypes associated with HLA variants. *Sci Transl Med*. 2017;9(389):eaai8708.
6. Bowness P. HLA-B27. *Annu Rev Immunol*. 2015;33:29–48.
7. Hollenbach JA, Oksenberg JR. The immunogenetics of multiple sclerosis: a comprehensive review. *J Autoimmun*. 2015;64:13–25.
8. Naranbhai V, Viard M, Dean M, Groha S, Braun DA, Labaki C, et al. HLA-A*03 and response to immune checkpoint blockade in cancer: an epidemiological biomarker study. *Lancet Oncol*. 2022;23(1):172–84.
9. Castro-Santos P, Rojas-Martinez A, Riancho JA, Lapunzina P, Flores C, Caracedo Á, et al. HLA-A*11:01 and HLA-C*04:01 are associated with severe COVID-19. *HLA*. 2023;102(6):731–9.
10. Hovhannisyán A, Madelian V, Avagyan S, Nazaretyan M, Hyussyan A, Sirunyan A, et al. HLA-C*04:01 affects HLA class I heterozygosity and predicted affinity to SARS-CoV-2 peptides, and in combination with age and sex of Armenian patients contributes to COVID-19 severity. *Front Immunol*. 2022;13: 769900.
11. Motzer RJ, Robbins PB, Powles T, Albiges L, Haanen JB, Larkin J, et al. Avelumab plus axitinib versus sunitinib in advanced renal cell carcinoma: biomarker analysis of the phase 3 JAVELIN Renal 101 trial. *Nat Med*. 2020;26(11):1733–41.
12. Kerr J, Anderson C, Lippman SM. Physical activity, sedentary behaviour, diet, and cancer: an update and emerging new evidence. *Lancet Oncol*. 2017;18(8):e457–71.
13. Collins R. What makes UK Biobank special? *Lancet*. 2012;379(9822):1173–4.
14. Palmer LJ. UK Biobank: bank on it. *Lancet*. 2007;369(9578):1980–2.
15. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med*. 2015;12(3): e1001779.
16. UK Biobank. UK Biobank data showcase.
17. Allen I, Hassan H, Sofianopoulou E, Eccles D, Turnbull C, Tischkowitz M, et al. Risk of developing a second primary cancer in male breast cancer survivors: a systematic review and meta-analysis. *Br J Cancer*. 2022;127(9):1660–9.
18. Lj O. Multiple malignant neoplasms. *JAMA*. 1921;76(20):1329–33.
19. Macq G, Silversmit G, Verdoodt F, Van Eycken L. The epidemiology of multiple primary cancers in Belgium (2004–2017): incidence, proportion, risk, stage and impact on relative survival estimates. *BMC Cancer*. 2023;23(1):349.
20. Pandurengan RK, Dumont AG, Araujo DM, Ludwig JA, Ravi V, Patel S, et al. Survival of patients with multiple primary malignancies: a study of 783 patients with gastrointestinal stromal tumor. *Ann Oncol*. 2010;21(10):2107–11.
21. Sung H, Siegel RL, Hyun N, Miller KD, Yabroff KR, Jemal A. Subsequent primary cancer risk among 5-year survivors of adolescent and young adult cancers. *J Natl Cancer Inst*. 2022;114(8):1095–108.
22. Tanjak P, Suktitipat B, Vorasan N, Juengwiwattanakit P, Thiangtrong B, Songjang C, et al. Risks and cancer associations of metachronous and synchronous multiple primary cancers: a 25-year retrospective study. *BMC Cancer*. 2021;21(1):1045.
23. Behrens M, Trejo T, Luthra H, Griffiths M, David CS, Taneja V. Mechanism by which HLA-DR4 regulates sex-bias of arthritis in humanized mice. *J Autoimmun*. 2010;35(1):1–9.
24. Flesch BK, König J, Frommer L, Hansen MP, Kahaly GJ. Sex alters the MHC class I HLA-A association with polyglandular autoimmunity. *J Clin Endocrinol Metab*. 2019;104(5):1680–6.
25. Irizar H, Muñoz-Culla M, Zuriarrain O, Goyenechea E, Castillo-Triviño T, Prada A, et al. HLA-DRB1*15:01 and multiple sclerosis: a female association? *Mult Scler*. 2012;18(5):569–77.
26. Jo YG, Ortiz-Fernández L, Coit P, Yilmaz V, Yentür SP, Alibaz-Oner F, et al. Sex-specific analysis in Behçet's disease reveals higher genetic risk in male patients. *J Autoimmun*. 2022;132:102882.
27. Nøst TH, Alcalá K, Urbarova I, Byrne KS, Guida F, Sandanger TM, et al. Systemic inflammation markers and cancer incidence in the UK Biobank. *Eur J Epidemiol*. 2021;36(8):841–8.
28. Cupp MA, Cariolou M, Tzoulaki I, Aune D, Evangelou E, Berlanga-Taylor AJ. Neutrophil to lymphocyte ratio and cancer prognosis: an umbrella review of systematic reviews and meta-analyses of observational studies. *BMC Med*. 2020;18(1):360.
29. Åberg AM, Bergström SH, Thysell E, Tjon-Kon-Fat LA, Nilsson JA, Widmark A, et al. High monocyte count and expression of S100A9 and S100A12 in peripheral blood mononuclear cells are associated with poor outcome in patients with metastatic prostate cancer. *Cancers (Basel)*. 2021;13(10):2424.
30. Feng F, Zheng G, Wang Q, Liu S, Liu Z, Xu G, et al. Low lymphocyte count and high monocyte count predicts poor prognosis of gastric cancer. *BMC Gastroenterol*. 2018;18(1):148.
31. Shigeta K, Kosaka T, Kitano S, Yasumizu Y, Miyazaki Y, Mizuno R, et al. High absolute monocyte count predicts poor clinical outcome in patients with castration-resistant prostate cancer treated with docetaxel chemotherapy. *Ann Surg Oncol*. 2016;23(12):4115–22.
32. Zhang LN, Xiao W, OuYang PY, You K, Zeng ZF, Ding PR, et al. The prognostic impact of preoperative blood monocyte count in pathological T3N0M0 rectal cancer without neoadjuvant chemoradiotherapy. *Tumour Biol*. 2015;36(10):8213–9.
33. Prizment AE, Anderson KE, Visvanathan K, Folsom AR. Inverse association of eosinophil count with colorectal cancer incidence: atherosclerosis risk in communities study. *Cancer Epidemiol Biomarkers Prev*. 2011;20(9):1861–4.
34. Wang JH, Rabkin CS, Engels EA, Song M. Associations between eosinophils and cancer risk in the UK Biobank. *Int J Cancer*. 2024;155(3):486–92.
35. Mentella MC, Scaldaferrri F, Ricci C, Gasbarrini A, Miggiano GAD. Cancer and Mediterranean diet: a review. *Nutrients*. 2019;11(9):2059.
36. García-Montero C, Fraile-Martínez O, Gómez-Lahoz AM, Pekarek L, Castellanos AJ, Nogueiras-Fraguas F, et al. Nutritional components in Western diet versus Mediterranean diet at the gut microbiota-immune system interplay Implications for health and disease. *Nutrients*. 2021;13(2):699.
37. Te Morenga L, Mallard S, Mann J. Dietary sugars and body weight: systematic review and meta-analyses of randomised controlled trials and cohort studies. *BMJ*. 2012;346: e7492.
38. Yan RR, Chan CB, Louie JCY. Current WHO recommendation to reduce free sugar intake from all sources to below 10% of daily energy intake for supporting overall health is not well supported by available evidence. *Am J Clin Nutr*. 2022;116(1):15–39.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.