Risk of Second Primary Cancers After a Diagnosis of First Primary Cancer: A Pan-Cancer Analysis and Mendelian Randomization Study

Xiaohao Ruan†, Da Huang†, Yongle Zhan², Jingyi Huang¹, Jinlun Huang¹, Ada Tsui-Lin NG², James Hok-Leung TSU², Rong Na*.

1. Department of Urology, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200025, China.

2. Division of Urology, Department of Surgery, Queen Mary Hospital, The University of Hong Kong, Hong Kong, China.

† Xiaohao Ruan and Da Huang contributed equally to this work.

* Corresponding Author.

Rong Na,
Division of Urology, Department of Surgery, Queen Mary Hospital, The University of Hong Kong, Hong Kong, China

Tel: +852-22554310; Email address: narong.hs@gmail.com.

No. of Words:

Abstract: 238 words

Text: 2956 words
Abstract

Background: The risk of second primary cancers (SPC) is increasing after the first primary cancers (FPC) are diagnosed and treated. The underlying causal relationship remains unclear.

Methods: We conducted a pan-cancer association (26 cancers) study in the Surveillance, Epidemiology, and End Results (SEER) database. The standardized incidence ratio (SIR) was estimated as the risk of SPCs in cancer survivors based on the incidence in the general population. Furthermore, the causal effect was evaluated by two-sample Mendelian Randomization (MR, 13 FPCs) in the UK Biobank (UKB, n=459,136) and robust analysis (radial MR and Causal Analysis Using Summary Effect estimates, CAUSE).

Results: We found 11 significant cross-correlations among different cancers after harmonizing SIR and MR results. Whereas only 4 of them were confirmed by MR to have a robust causal relationship. In particular, patients initially diagnosed with oral pharyngeal cancer would have an increased risk of non-Hodgkin lymphoma (SIR_{SEER}=1.18, 95% Confidence Interval [CI]:1.05-1.31, OR_{radial-MR}=1.21, 95%CI:1.13-1.30, P=6.00×10^{-3}; OR_{cause}=1.17, 95%CI:1.05-1.31, P=8.90×10^{-3}).

Meanwhile, ovary cancer was identified to be a risk factor for soft tissue cancer (SIR_{SEER}=1.72, 95% Confidence Interval [CI]:1.08-2.60, OR_{radial-MR}=1.39, 95%CI:1.22-1.58, P=1.07×10^{-3}; OR_{cause}=1.36, 95%CI:1.16-1.58, P=0.01). And kidney cancer was likely to cause the development of lung cancer (SIR_{SEER}=1.28,
95% Confidence Interval [CI]: 1.22-1.35, OR_{radial-MR} = 1.17, 95% CI: 1.08-1.27, $P = 6.60 \times 10^{-3}$; OR_{cause} = 1.16, 95% CI: 1.02-1.31, $P = 0.05$) and myeloma (SIR_{SEER} = 1.54, 95% Confidence Interval [CI]: 1.33-1.78, OR_{radial-MR} = 1.24, 95% CI: 1.21-2.45, $P = 0.02$; OR_{cause} = 1.49, 95% CI: 1.04-2.34, $P = 0.02$).

**Conclusions:** A certain type of primary cancer may cause another second primary cancer, and the profound mechanisms need to be studied in the future.

**Funding:** This work was supported by grants from National Natural Science Foundation of China (Grant No. 81972645), Innovative research team of high-level local universities in Shanghai, Shanghai Youth Talent Support Program, intramural grant of The University of Hong Kong to Dr. Rong Na, and Shanghai Sailing Program (22YF1440500) to Dr. Da Huang.

**Keywords:** First primary cancer; Mendelian Randomization study; Pan-cancer analysis; Second primary cancer.
Introduction

Cancer incidence is rapidly growing worldwide in the past decades. The reasons are complex including the aging of the population, the application of screening, environmental risk factors and genetic risk factors (Sung et al., 2021). In 2020, there were an estimated 19.3 million new cases and 10.0 million deaths of different types of cancers globally (Sung et al., 2021). In the US, there would be about 1.9 million new cases of cancers in 2022, and the most common type of cancers in the male and the female were prostate cancer (27%) and breast cancer (31%), respectively (Siegel et al., 2022). Despite the rapidly increased incidence of cancers, the survival of (most types of) cancers in the US has vastly improved since the mid-1970s with medical advances and technical developments (Siegel et al., 2022). For example, the 5-year relative survival rates of prostate cancer, melanoma, and female breast cancer, during 2011-2017 in the US were 98%, 93%, and 90%, respectively (Siegel et al., 2022). Such disparity may lead to an increase in the prevalence and the tumor burden in US society. More importantly, prolonged survival makes it possible for individuals to be diagnosed with a second primary cancer (SPC) after the first primary cancer (FPC) during the follow-up.

According to the Italian (1976-2010) ("Italian cancer figures, report 2013: Multiple tumours," 2013), the Swiss (1981-2009) (Feller et al., 2020), and the Swedish (1990-2015) (Zheng et al., 2020) cancer registration data, increased risks of SPCs were observed in many types of cancer as the FPCs. Patients with oral cavity & pharynx, larynx, and esophagus as FPC were found to have a significantly elevated risk of any
SPCs in both Italy and Switzerland (Feller et al., 2020; "Italian cancer figures, report 2013: Multiple tumours," 2013). In Sweden, liver cancers, as well as nasal and oral cancers, were found to be associated with SPCs (Zheng et al., 2020). In addition, several studies focused on certain cancer also suggested a potential relationship between FPC and SPC (Chattopadhyay et al., 2018). For example, increased FPC risk of colorectal cancer, kidney cancer, and melanoma were observed following the diagnosis of non-Hodgkin lymphoma. Despite the strong association observed in these studies, whether there is any underlying causal relationship is unknown, or the association observed is due to the potential confounders such as aging.

In the present study, our objectives are to perform a pan-cancer association study in the Surveillance, Epidemiology, and End Results (SEER) and to interpret the underlying causal relationship via Mendelian Randomization approaches using genetic variants in a large population cohort (UK Biobank, UKB). In addition, the study may also help us understand critical questions in clinical practice about who should be more careful of the second primaries, and who should take precision screenings against certain cancers in addition to the regular follow-up evaluations.

**Results**

We set out to identify the observational association between FPCs and SPCs among 22 solid tumors and 4 hematological malignancies based on the SEER database (Figure 1). A total of 264 significant standardized incidence ratio (SIR)
existed among them when compared to a standard population (Table S1-17).

Hierarchical clustering analysis (heatmap) is shown in Figure 2 which illustrates the comparison between the cancer incidence in patients with a certain type of FPC and the incidence in the population average level in the SEER dataset. Second primary thyroid cancer, small bowel cancer, or cancers of endocrine system were strongly and significantly associated with most of the FPCs. Cancers of the digestive system, cervix uteri, as well as lung cancer were clustered together. They are closely correlated with an increased risk of several types of cancers such as bladder cancer, kidney cancer, etc. Our subgroup analysis suggested that patients with prostate cancer who received radiation therapy were at an increased risk of being diagnosed with another type of primary cancer during the follow-up including small intestine, soft tissue, and leukemia, compared with those without such therapy (Table S13). And SPC risk after breast cancer was inconsistent among males and females. Men with breast cancer had a higher risk for thyroid cancer and prostate cancer (SIRSEER=1.30; 95% CI, 1.13-1.49; SIRSEER=2.33, 1.12-4.29; Table S10), but with no significant risk change for other cancers.

Details of the included exposure-associated SNPs in European ancestry were shown in Table 1 and Table S18. The number of SNPs ranged from 5 (esophagus) to 104 (female breast), and the proportion of variance explained by SNPs (R²) ranged from 0.10% (thyroid) to 8.46% (bladder). F-statistics for all 13 cancers exceeded 10, suggesting no weak instrument bias here. However, some problems, including too few
SNPs, no related GWAS and disease heterogeneity, affected the comprehensive MR analysis (Table 1). Due to the low incidence of cancers in UKB, the power to detect a significant effective size (0.8/1.2) was relatively low, except for colorectal cancer, lung cancer, female breast cancer and prostate cancer (Table S19).

Results from MR analyses are presented in Table S20-33. A total of 23 significant association was detected (16 positive causality and 7 negative causality). Concordant significant results and unconcordant results between MR and the SEER SIR analyses are shown in Table 2 and Table S20, respectively. The concordant results suggested that patients diagnosed with primary oral and pharynx cancer would cause a significantly increased risk of second primary non-Hodgkin lymphoma (SIR_{SEER}=1.18, 95% CI: 1.05-1.31; IVW-MR $P=8.96 \times 10^{-4}$). After a primary diagnosis of pancreatic cancer, SPC risks were increased for small intestine (SIR_{SEER}=4.37, 95% CI: 2.85-6.40; MR-Egger $P=0.04$). It also indicated that female patients initially diagnosed with melanoma would cause a mild but significantly increased risk of cancers in the breast (SIR_{SEER}=1.17, 95% CI: 1.12-1.23; IVW-MR $P=0.04$), as well as ovary cancer on soft tissue related cancer (SIR_{SEER}=1.72, 95% CI: 1.08-2.60; IVW-MR $P=8.39 \times 10^{-5}$). The greatest number of casual relationships were observed in kidney cancer as FPC. A primary kidney cancer might cause an elevated risk of cancers of lung and bronchus (SIR_{SEER}=1.28, 95% CI: 1.23-1.45; IVW-MR $P=0.01$), non-Hodgkin lymphoma (SIR_{SEER}=1.19, 95% CI: 1.08-1.31; IVW-MR $P=3.64 \times 10^{-3}$), myeloma (SIR_{SEER}=1.54, 95% CI: 1.33-1.78; IVW-MR $P=3.94 \times 10^{-3}$). Meanwhile, some primary cancer site
might give protective effect against another cancer (for instance, lung cancer vs. female breast cancer, Table 2). Scatter plot, Funnel plot, forest plot and leave-one-out analysis showed single SNP effective size in Figure S1-4, respectively.

More conservative analyses were performed to further confirm these causal relationships. We applied 2 outlier-detected methods with modified second order weights (radial IVW and radial MR-Egger) and CAUSE (Causal Analysis Using Summary Effect estimates) to each pair of phenotypes, with the rationale that robust relationships would exhibit consistent and statistically significant results across different methods. Additionally, CAUSE is the only method capable of distinguishing causality from both correlated and uncorrelated pleiotropy. The relationship with at least 2 significant results was treated as a robust causality (Table 3). We found consistent evidence for a causal effect of oral and pharynx cancer on non-Hodgkin lymphoma ($P_{\text{radial-IVW}} = 6.00 \times 10^{-3}$, $P_{\text{cause}} = 8.90 \times 10^{-3}$), ovary cancer on soft tissue cancer ($P_{\text{radial-IVW}} = 1.07 \times 10^{-3}$, $P_{\text{cause}} = 0.01$), kidney cancer on lung and myeloma ($P_{\text{radial-IVW}} = 6.60 \times 10^{-3}$, $P_{\text{cause}} = 0.05$; $P_{\text{radial-IVW}} = 0.02$, $P_{\text{cause}} = 0.02$).

Discussion

With the expanded life expectancy and the prolonged survival of cancers, the incidence of SPCs has been rapidly growing in the past decades (Copur & Manapuram, 2019). Genetic factors or shared environmental factors are probably the major causes. In the previous association studies, individuals with a certain type of primary
malignancy would have an increased risk of another malignancy (Feller et al., 2020; "Italian cancer figures, report 2013: Multiple tumours," 2013; Zheng et al., 2020). However, whether there are any causal effects within the associations is unclear. In the present study, via the association study based on the SEER database and the MR approach using the UKB genetic dataset, we were able to perform this comprehensive investigation across 26 different types of cancers. 13 out of 26 types of cancers with adequate GWAS data were able to be further investigated using MR analysis. We found that numbers of primary malignancies were associated with an increased risk of a second primary malignancy, however, only a small part of the associations would have a causal relationship (Table 2).

Many significant findings were observed in the SEER SIR analysis. SEER is one of the largest cancer registration-based datasets making itself the most proper data source to answer the study objectives; however, several advantages of this database and limitations of the results should be noted. First, the results from the SEER SIR analysis were associations rather than causal inferences. Many factors may influence the results of the associations. For example, the confounder of screening effects may exist. Patients diagnosed with primary cancer might have more frequent healthcare visits compared to those who did not have any cancers. Therefore, some indolent cancers such as thyroid cancer and low-risk prostate cancer could have been over-diagnosed due to the screening effects. Second, it is a cancer registration-based cohort rather than a population cohort, the standardized incidences calculated from the SEER database...
may not represent the situation in the general population. Third, treatment preferences and follow-up strategies may vary in different locations or institutions, which would also affect the occurrence/detection of the second primary cancer. For example, radiation therapy may increase the risk of cancer in nearby organs. However, treatment and follow-up information are not completed in the SEER database due to the natural design of the cohort. More importantly, lifestyles, comorbidities, and environmental factors were not included in the SEER database. These are important confounders of the associations between the first primary cancers and the second primary cancers.

The MR approaches in the current study revealed some interesting findings, but several non-concordant results between MR analyses and SEER SIR were also observed (Table S20). It does not indicate that the causal relationship does not exist. Some factors, such as the period of follow-up, may conceal and cause false negatives in the association study (SEER SIR). For example, pancreatic cancer might cause an increased risk of cancers in esophagus, colon and rectum, etc. based on the MR analysis in our study; however, no association was observed in SEER SIR analysis. The short and poor survival of pancreatic cancer could be the most critical reason for the failure of finding a positive association in the population data --- simply did not have enough time of follow-up to observe the outcomes. Therefore, the interpretation of these results should be more careful at this stage.

The lack of GWAS findings would be a major limitation of the MR approach for some diseases as in the present study. The MR approach may only represent part of...
the biological effects in the causal pathway between the exposure and the outcome. A
final causal inference should always be established based on biological mechanisms.
From the angle of organ location, some cancers (ovary cancer and soft tissue sarcoma) might share the same tumor-related or tumor-developing environment. Besides the outside therapeutic settings (radio- and chemotherapy, immuno-suppressive agents) and individual factors (smoking, hormone level, certain occupational settings, HIV or HPV infections, and family histories), FPC might also influence the iatrogenic immune by suppressing antitumor defense mechanisms via inflammation or other meditating effects (Shalapour & Karin, 2019). For instance, the increased risk of renal cancer, and non-Hodgkin lymphoma were observed in immunosuppressed patients in Denmark and Sweden (Hortlund et al., 2017). The immune factor might be the inner relation between FPC and SPC. In our assumption, the question of mechanism should be answered via cross-trait GWAS meta-analyses, searching for shared genetic architecture under high heritability or potential mediation factors with comprehensive database and analyses, functional experiments (tissue- or cell-specific findings) and final validation in a cohort of comorbidity patients.

Finally, besides the limitations mentioned, the relatively small number of cases of some diseases in UKB may lower our statistical power. As UKB is a population-based prospective cohort, the relatively short follow-up period may not allow us to observe enough events (multiple cancers) at this stage. And it is expected to be independently replicated in another dataset. Regardless, these findings progress our
understanding of the relationship underlying both FPCs and SPCs, and potentially provide points of exploration and intervention that may reduce second primary cancers.

Conclusion

Patients who were diagnosed with a certain type of primary cancer may cause another type of primary cancer, especially pharynx cancer on non-Hodgkin lymphoma, ovary cancer on soft tissue cancer, kidney cancer on lung and myeloma. The profound mechanisms need to be studied in the future.

Methods

Study populations

SEER Program 18 Registry database was obtained which covered 27.8% of the total population in the United States ("Number of Persons by Race and Hispanic Ethnicity for SEER Participants - SEER Registries."). The SEER program is the largest cancer incidence dataset in the United States based on population cancer registration. Based on the ICD-10 code, we identified adult patients (age ≥20 years) diagnosed with an FPC between 2000 and 2016, including 22 types of solid-tumor sites (oral cavity and pharynx, esophagus, stomach, small intestine, colon and rectum, liver, gallbladder, pancreas, larynx, lung and bronchus, bones and joints, soft tissue including heart, melanoma of the skin, female breast, male breast, cervix uteri, ovary,
prostate, bladder, kidney, renal pelvis and ureter, brain, thyroid) and 4 types of hematological malignancies (Hodgkin lymphoma, non-Hodgkin lymphoma, myeloma, leukemia). Patients with diagnosis by autopsy or mentioned in the death certificate only were excluded.

The UKB project is a prospective cohort study collecting phenotypic and genotypic data from ~500,000 individuals from across the United Kingdom (median follow-up time was ~14 years). The participants aged between 40 and 69 at recruitment (Bycroft et al., 2018). In the present study, a total of 459,156 participants with European Ancestry from UKB (release V3) with GWAS genotyping array data and imputation information were obtained and included in the MR analysis. Disease phenotypes in UKB were also defined using the ICD-10 code. Non-Caucasian patients were not included in the present study in SEER or UKB due to the small number of subjects in UKB, which made it hard to make the causal inference.

Written informed consent was obtained from all the participants from SEER or UKB according to the established standard of the studies. The current study design was reviewed by the Institutional Review Board at Ruijin Hospital (Shanghai, China).

Genotyping and Quality Control

GWAS genotyping array data with imputation and QC from UK Biobank release V3 was obtained (Bycroft et al., 2018). Briefly, a total of 488,377 participants (after quality control, QC) were genotyped in UKB using two similar genotyping arrays,
the UK Biobank Lung Exome Variant Evaluation (UK BiLEVE with 807,411 markers, n=49,950) and the Applied Biosystems UK BiLEVE Axiom Array by Affymetrix (825,927 markers, n=438,427). These two arrays share 95% of the markers. Individuals were excluded if: (a) ancestry testing using principal component analysis (PCA) to evaluate the potential conflicts between self-reported ethnicity/race and the genetic ethnicity/race; (b) Poor call rate at the genotyping stage (n=968, 0.2%); (c) Mismatched results between self-reported gender and genetic gender (n=652, 0.13%). The genotype concordance rate was reported as >99.0% (Bycroft et al., 2018). A total of 93,095,623 autosomal SNPs were identified in 487,442 individuals (Bycroft et al., 2018).

Mendelian Randomization

The conceptual framework was illustrated in Figure 1 and the MR study was reported in accordance with the STROBE-MR guideline (Skrivankova et al., 2021). Two-sample MR analyses were performed to evaluate the causality between exposures (a certain primary cancer, FPC) and outcomes (another primary cancer, SPC). We used previously identified disease risk-associated SNPs from the GWAS Catalog database (https://www.ebi.ac.uk/gwas/) (Buniello et al., 2019). SNPs selection was based on the following criteria: (1) SNPs were from a single GWAS that identified the largest number of risk-associated SNPs and were conducted in European ancestry; (2) SNPs reached genome-wide significant level ($P<5 \times 10^{-8}$); (3)
SNPs are independent of each other in terms of linkage disequilibrium (LD, \( r^2 < 0.001 \)) and distance (>10,000 kb).

MR analyses derive valid estimates where the following assumptions are met: (i) the SNPs are correlated with FPC, (ii) the SNPs affect SPC risk only through their effects on FPC and (iii) the SNPs are independent of any confounding factors for the association between FPC and SPC. For assumption (i), the strength of each instrument was measured using the F statistic and the proportion of the explained variance (R\(^2\)), which was considered to be sufficient if the corresponding F-statistic is >10. For assumption (ii) and (iii), we searched the PhenoScanner database (available at http://www.phenoscanner.medschl.cam.ac.uk/phenoscanner) (Kamat et al., 2019; Staley et al., 2016) to examine whether SNPs were significantly associated with established risk factors for certain cancers, including BMI, smoking, alcohol intake and physical inactivity, and excluded those at \( P < 1.0 \times 10^{-5} \). Statistical power calculations were performed using an online tool available at https://shiny.cnsgenomics.com/mRnd/(Brion et al., 2013). The statistical power was to capture an OR of 1.20 or 0.80 per a standard deviation (SD) change in the cancer risk.

Inverse-variance weighted MR (IVW-MR) and MR-Egger were used in the MR analyses (Burgess et al., 2019; Burgess & Thompson, 2017; Davies et al., 2018; Hemani et al., 2018). MR would be performed based on at least 4 SNPs. Briefly, these two methods are the most used MR methods to infer a causal relationship. IVW-MR is
based on a random effect model and is the most efficient with the greatest statistical
power (Hemani et al., 2018). Potential bias as horizontal pleiotropy was evaluated and
adjusted via MR-Egger (Burgess et al., 2019; Burgess & Thompson, 2017; Davies et al.,
2018; Hemani et al., 2018). The causal inference was interpreted via IVW-MR results if
the horizontal pleiotropic effect was not significant; otherwise, based on MR-Egger. A
causal relationship will only be interpreted when a significant MR result was observed
from MR analyses together with a significant association based on the SEER database.
A series of sensitivity and robust analyses including leave-one-out, Radial
MR (Bowden et al., 2019; Bowden et al., 2018) and CAUSE (Morrison et al., 2020)
methods would be performed in case of concordant significant results. MR analyses
were performed using the R package “TwoSampleMR”, “MendelianRandomization”,
“RadialMR” and “cause”.

Statistical Analysis

The multiple primary standardized incidence ratios (MP-SIR) were defined as the
observed incidence of a second malignancy among cases previously diagnosed with a
certain type of cancer divided by the expected incidence based on the SEER referent
population (the SEER18 2000-2016 referent rate file). All the standardized incidence
observed/expected (O/E) ratios (SIR) and the corresponding 95% confidence intervals
(95% CI) were derived using the “MP-SIR” session of SEER*Stat (version
8.3.8)("Multiple Primary - Standardized Incidence Ratios - SEER*Stat."). Only the
A non-Hispanic white population was included in the present study. We restricted a minimum 2-month latency period between the first and second primary diagnosis (Davis et al., 2014). Subgroup analyses were performed after stratifying by radiation therapy (received or not). All statistical analyses were performed using SEER*STAT and R software (4.1.2) (Team, 2021). A type I error of 0.05 (two-sided) was used to define statistical significance. Multiplicity effects were only considered during the selection of SNPs.
Reference


https://doi.org/10.3322/caac.21660

https://www.R-project.org/.


Data Availability

Data used in this research are publicly available to qualified researchers on application to GWAS catalog (https://www.ebi.ac.uk/gwas/), the SEER database (https://seer.cancer.gov/) and UK Biobank (www.ukbiobank.ac.uk).

Author Contributions

R Na conceived and designed the study. XH Ruan, Da Huang processed and analyzed the microarray data, performed different components of analysis and data interpretations. YL Zhan provided technical support. JY Huang and JL Huang provided material support. ATL Ng and JHL Tsu provided administrative support. XH Ruan, D Huang and R Na wrote the manuscript. R Na made critical revision of the manuscript for important intellectual content. All authors read and approved the final manuscript.

Competing interest
The authors declare that there are no competing interests.

**Acknowledgement**

The current study contains a part of the capstone thesis by Rong Na for Master of Public Health at Johns Hopkins University School of Public Health (the results were updated based on the updated datasets); otherwise, the study content has never been published elsewhere. We would like to thank for the supervision from Professor Jianfeng Xu from NorthShore University Health System, IL USA, Professor William B. Issacs and Professor Bruce J. Trock from Johns Hopkins University, during Rong Na’s capstone projects. We thank SEER program to approve our protocol and provide the custom datasets. We also thank the UK Biobank for access to the data (Project Number: 66813).

**Funding information**

This work was in supported by grants from National Natural Science Foundation of China (Grant No. 81972645), Innovative research team of high-level local universities in Shanghai, Shanghai Youth Talent Support Program, intramural grant of The University of Hong Kong to Dr. Rong Na, and Shanghai Sailing Program (22YF1440500) to Dr. Da Huang. All the funders had no role in study design, data collection, data analysis, interpretation, and writing of the report.
Figure legend

Figure 1. Study design and work flowchart.

Abbreviation: SIR, standardized incidence ratio; GWAS, genome-wide significant association.
Figure 2. Heat-map of SIR for First primary cancers (FPC, horizontal, cancer listed in the bottom) and second primary cancers (SPC, vertical, cancer listed on the right) in the SEER 18 registry (2000-2016).

The SIR scale is shown in the right corner. Only significant associations (95% confidence intervals not overlapping with 1.00) were included; SIRs of the insignificant or concordant associations were assigned as 1. Abbreviations: SIR, standardized incidence ratios; FPC, first primary cancers; SPC, second primary cancers.
Table 1. Summary of the cancer-specific instrument variables used in this study (European ancestry).

<table>
<thead>
<tr>
<th>Cancer type</th>
<th>GWAS Source</th>
<th>PMID</th>
<th>Number of SNPs(^a)</th>
<th>Cases</th>
<th>Total population</th>
<th>Variance explained, (R^2, %)</th>
<th>F-statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oral and pharynx</td>
<td>GCST003857</td>
<td>27749845</td>
<td>7</td>
<td>6,034</td>
<td>12,619</td>
<td>2.22</td>
<td>283.68</td>
</tr>
<tr>
<td>Larynx</td>
<td>GCST010285</td>
<td>32276964</td>
<td>1(^b)</td>
<td>394</td>
<td>4,887</td>
<td>0.28</td>
<td>13.72</td>
</tr>
<tr>
<td>Esophagus</td>
<td>GCST003740</td>
<td>27527254</td>
<td>5</td>
<td>10,279</td>
<td>27,438</td>
<td>0.72</td>
<td>198.94</td>
</tr>
<tr>
<td>Stomach</td>
<td>GCST002990</td>
<td>26098866</td>
<td>1(^b)</td>
<td>2,043</td>
<td>202,533</td>
<td>0.02</td>
<td>34.44</td>
</tr>
<tr>
<td>Colon and rectum</td>
<td>GCST003017</td>
<td>26151821</td>
<td>8</td>
<td>18,299</td>
<td>37,955</td>
<td>0.89</td>
<td>340.75</td>
</tr>
<tr>
<td>Pancreas</td>
<td>GCST005434</td>
<td>29422604</td>
<td>14</td>
<td>9,040</td>
<td>21,536</td>
<td>4.28</td>
<td>962.28</td>
</tr>
<tr>
<td>Melanoma</td>
<td>GCST004142</td>
<td>28212542</td>
<td>17</td>
<td>6,628</td>
<td>293,193</td>
<td>0.29</td>
<td>852.68</td>
</tr>
<tr>
<td>Lung</td>
<td>GCST004748</td>
<td>28604730</td>
<td>7</td>
<td>29,266</td>
<td>85,716</td>
<td>0.55</td>
<td>474.00</td>
</tr>
<tr>
<td>Male Breast</td>
<td>GCST011526</td>
<td>32785646</td>
<td>2(^b)</td>
<td>2,190</td>
<td>6,836</td>
<td>1.27</td>
<td>87.91</td>
</tr>
<tr>
<td>Female Breast</td>
<td>GCST004988</td>
<td>29059683</td>
<td>104</td>
<td>122,977</td>
<td>228,951</td>
<td>4.95</td>
<td>11917.81</td>
</tr>
<tr>
<td>Cervix uteri</td>
<td>GCST004833</td>
<td>28806749</td>
<td>1(^b)</td>
<td>2,866</td>
<td>9,347</td>
<td>0.45</td>
<td>42.24</td>
</tr>
<tr>
<td>Ovary</td>
<td>GCST002748</td>
<td>25581431</td>
<td>9</td>
<td>18,530</td>
<td>69,745</td>
<td>1.16</td>
<td>818.42</td>
</tr>
<tr>
<td>Prostate</td>
<td>GCST006085</td>
<td>29892016</td>
<td>71</td>
<td>79,148</td>
<td>140,254</td>
<td>5.47</td>
<td>8111.66</td>
</tr>
<tr>
<td>Bladder</td>
<td>GCST002240</td>
<td>24163127</td>
<td>7</td>
<td>2,305</td>
<td>6,206</td>
<td>8.46</td>
<td>572.81</td>
</tr>
<tr>
<td>Kidney</td>
<td>GCST004710</td>
<td>28598434</td>
<td>8</td>
<td>10,784</td>
<td>31,190</td>
<td>1.11</td>
<td>349.99</td>
</tr>
<tr>
<td>Phenotype</td>
<td>Study ID</td>
<td>Study ID Description</td>
<td>$n$</td>
<td>Total SNPs</td>
<td>Total Genomic DNA</td>
<td>PRS</td>
<td>PRS Score</td>
</tr>
<tr>
<td>------------</td>
<td>-------------</td>
<td>----------------------</td>
<td>-----</td>
<td>------------</td>
<td>------------------</td>
<td>-----</td>
<td>-----------</td>
</tr>
<tr>
<td>Thyroid</td>
<td>GCST004144</td>
<td></td>
<td>6</td>
<td>3,001</td>
<td>290,551</td>
<td>0.10</td>
<td>290.83</td>
</tr>
<tr>
<td>Myeloma</td>
<td>GCST004483</td>
<td></td>
<td>13</td>
<td>9,866</td>
<td>249,054</td>
<td>0.31</td>
<td>774.24</td>
</tr>
</tbody>
</table>

- Number of SNPs included in the final calculation of PRS in our study; not necessarily the total number of SNPs from the source due to the filtering steps discussed in the main text and germline data availability; $^b$ MR would be performed based on at least 4 SNPs.

Note: See Additional file and supplementary materials: Table S18 for the list of SNPs included in the final calculation for each phenotype.

- No study found in European: bone and joint, brain, liver, small intestine, gallbladder, renal pelvis and ureter;
- Heterogeneity: lymphoma, leukemia;
- Too few SNPs: larynx, stomach, male breast cancer.
Table 2. Concordant causality between Mendelian randomization results and SEER analysis.

<table>
<thead>
<tr>
<th>Cancer type (first primary cancer)</th>
<th>Positive Causality</th>
<th>Negative Causality</th>
<th>No Causality (statistical power ≥ 0.80)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oral and pharynx</td>
<td>Non-hodgkin lymphoma</td>
<td>-</td>
<td>Female breast, Prostate</td>
</tr>
<tr>
<td>Esophagus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Colon and rectum</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Pancreas</td>
<td>Small intestin</td>
<td>-</td>
<td>Melanoma</td>
</tr>
<tr>
<td>Melanoma</td>
<td>Female breast</td>
<td>-</td>
<td>Colon and Rectum</td>
</tr>
<tr>
<td>Lung</td>
<td>-</td>
<td>Female Breast</td>
<td>-</td>
</tr>
<tr>
<td>Female Breast</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Ovary</td>
<td>Soft tissue</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Prostate</td>
<td>-</td>
<td>Colon and Rectum</td>
<td>Non-hodgkin lymphoma</td>
</tr>
<tr>
<td>Bladder</td>
<td>-</td>
<td>-</td>
<td>Female breast, Non-hodgkin lymphoma</td>
</tr>
<tr>
<td>Kidney</td>
<td>Lung and Bronchus, Melanoma, Non-hodgkin lymphoma, Myeloma</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Thyroid</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Myeloma</td>
<td>-</td>
<td>-</td>
<td>Lung and Bronchus</td>
</tr>
</tbody>
</table>

See Additional file: Table S20 for unconcordant causality result; Table S21-S33 for the details of Mendelian randomization results (IVW and MR-Egger).
<table>
<thead>
<tr>
<th>Exp-out</th>
<th>Radial IVW</th>
<th></th>
<th>Radial MR-Egger</th>
<th></th>
<th>CAUSE</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OR (95%CI)</td>
<td>P</td>
<td>OR (95%CI)</td>
<td>P</td>
<td>OR (95%CI)</td>
<td>P</td>
</tr>
<tr>
<td>Oropharynx-NHL</td>
<td>1.21 (1.13-1.30)</td>
<td>6.00×10⁻³</td>
<td>1.21 (1.13-1.30)</td>
<td>0.52</td>
<td>1.17 (1.05-1.31)</td>
<td>8.90×10⁻³</td>
</tr>
<tr>
<td>Pancreas-Intestin</td>
<td>1.05 (0.83-1.32)</td>
<td>0.69</td>
<td>2.39 (0.91-6.31)</td>
<td>0.10</td>
<td>1.03 (0.79-1.35)</td>
<td>1.00</td>
</tr>
<tr>
<td>Melanoma-Breast</td>
<td>1.08 (1.00-1.15)</td>
<td>0.06</td>
<td>1.05 (0.84-1.32)</td>
<td>0.69</td>
<td>1.04 (0.95-1.16)</td>
<td>0.75</td>
</tr>
<tr>
<td>Lung-Breast</td>
<td>0.86 (0.79-0.93)</td>
<td>9.76×10⁻³</td>
<td>0.77 (0.54-1.08)</td>
<td>0.19</td>
<td>0.89 (0.76-1.04)</td>
<td>0.23</td>
</tr>
<tr>
<td>Ovary-Soft</td>
<td>1.39 (1.22-1.58)</td>
<td>1.07×10⁻³</td>
<td>1.3 (0.74-2.29)</td>
<td>0.39</td>
<td>1.36 (1.16-1.58)</td>
<td>0.01</td>
</tr>
<tr>
<td>Prostate-CRC</td>
<td>1.00 (0.95-1.04)</td>
<td>0.94</td>
<td>0.88 (0.80-0.98)</td>
<td>0.02</td>
<td>0.99 (0.93-1.04)</td>
<td>0.99</td>
</tr>
<tr>
<td>Kidney-CRC</td>
<td>1.17 (1.08-1.27)</td>
<td>6.60×10⁻³</td>
<td>0.94 (0.62-1.43)</td>
<td>0.78</td>
<td>1.16 (1.02-1.31)</td>
<td>0.05</td>
</tr>
<tr>
<td>Kidney-Melanoma</td>
<td>1.33 (1.02-1.73)</td>
<td>0.04</td>
<td>0.56 (0.18-1.80)</td>
<td>0.37</td>
<td>1.25 (0.96-1.73)</td>
<td>0.51</td>
</tr>
<tr>
<td>Kidney-NHL</td>
<td>1.25 (1.11-1.40)</td>
<td>7.87×10⁻³</td>
<td>1.33 (0.72-2.46)</td>
<td>0.39</td>
<td>1.20 (0.99-1.43)</td>
<td>0.09</td>
</tr>
<tr>
<td>Kidney-Myeloma</td>
<td>1.72 (1.21-2.45)</td>
<td>0.02</td>
<td>0.43 (0.11-1.77)</td>
<td>0.29</td>
<td>1.49 (1.04-2.34)</td>
<td>0.02</td>
</tr>
<tr>
<td>Myeloma-Lung</td>
<td>0.92 (0.86-0.98)</td>
<td>0.02</td>
<td>1.09 (0.79-1.51)</td>
<td>0.61</td>
<td>0.93 (0.86-1.00)</td>
<td>0.21</td>
</tr>
</tbody>
</table>

Abbreviation: IVW, Inverse variance weighted; CAUSE, Causal Analysis Using Summary Effect estimates; OR, odds ratio; CI, confidence interval; exp, exposure; out, outcome; NHL, non-hodgkin lymphoma; CRC, colorectal cancer.