

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

3  
4     Xiaohao Ruan<sup>1†</sup>, Da Huang<sup>1†</sup>, Yongle Zhan<sup>2</sup>, Jingyi Huang<sup>1</sup>, Jinlun Huang<sup>1</sup>, Ada  
5     Tsui-Lin NG<sup>2</sup>, James Hok-Leung TSU<sup>2</sup>, Rong Na<sup>2\*</sup>.

6  
7     1. Department of Urology, Ruijin Hospital, Shanghai Jiao Tong University School of  
8     Medicine, Shanghai 200025, China.

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

11    <sup>†</sup> Xiaohao Ruan and Da Huang contributed equally to this work.

12    \***Corresponding Author.**

13    Rong Na,

14    Division of Urology, Department of Surgery, Queen Mary Hospital, The University of  
15    Hong Kong, Hong Kong, China

16    Tel: +852-22554310; Email address: [narong.hs@gmail.com](mailto:narong.hs@gmail.com).

17    No. of Words:

18    Abstract: 238 words

19    Text: 2956 words

20

21

22 **Abstract**

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

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

33 **Results:** We found 11 significant cross-correlations among different cancers after  
34 harmonizing SIR and MR results. Whereas only 4 of them were confirmed by MR to  
35 have a robust causal relationship. In particular, patients initially diagnosed with oral  
36 pharyngeal cancer would have an increased risk of non-Hodgkin lymphoma  
37 ( $SIR_{SEER}=1.18$ , 95%Confidence Interval [CI]:1.05-1.31,  $OR_{radial-MR}=1.21$ ,  
38 95%CI:1.13-1.30,  $P=6.00\times 10^{-3}$ ;  $OR_{cause}=1.17$ , 95%CI:1.05-1.31,  $P=8.90\times 10^{-3}$ ).  
39 Meanwhile, ovary cancer was identified to be a risk factor for soft tissue cancer  
40 ( $SIR_{SEER}=1.72$ , 95%Confidence Interval [CI]:1.08-2.60,  $OR_{radial-MR}=1.39$ ,  
41 95%CI:1.22-1.58,  $P=1.07\times 10^{-3}$ ;  $OR_{cause}=1.36$ , 95%CI:1.16-1.58,  $P=0.01$ ). And kidney  
42 cancer was likely to cause the development of lung cancer ( $SIR_{SEER}=1.28$ ,

43 95%Confidence Interval [CI]:1.22-1.35,  $OR_{\text{radial-MR}}=1.17$ , 95%CI:1.08-1.27,  
44  $P=6.60\times 10^{-3}$ ;  $OR_{\text{cause}}=1.16$ , 95%CI:1.02-1.31,  $P=0.05$ ) and myeloma ( $SIR_{\text{SEER}}=1.54$ ,  
45 95%Confidence Interval [CI]:1.33-1.78,  $OR_{\text{radial-MR}}=1.24$ , 95%CI:1.21-2.45,  $P=0.02$ ;  
46  $OR_{\text{cause}}=1.49$ , 95%CI:1.04-2.34,  $P=0.02$ ).

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

49 **Funding:** This work was in supported by grants from National Natural Science  
50 Foundation of China (Grant No. 81972645), Innovative research team of high-level  
51 local universities in Shanghai, Shanghai Youth Talent Support Program, intramural  
52 grant of The University of Hong Kong to Dr. Rong Na, and Shanghai Sailing Program  
53 (22YF1440500) to Dr. Da Huang.

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

56

## 57 **Introduction**

58 Cancer incidence is rapidly growing worldwide in the past decades. The reasons  
59 are complex including the aging of the population, the application of screening,  
60 environmental risk factors and genetic risk factors(Sung et al., 2021). In 2020, there  
61 were an estimated 19.3 million new cases and 10.0 million deaths of different types of  
62 cancers globally(Sung et al., 2021). In the US, there would be about 1.9 million new  
63 cases of cancers in 2022,and the most common type of cancers in the male and the  
64 female were prostate cancer (27%) and breast cancer (31%), respectively(Siegel et al.,  
65 2022). Despite the rapidly increased incidence of cancers, the survival of (most types of)  
66 cancers in the US has vastly improved since the mid-1970s with medical advances and  
67 technical developments(Siegel et al., 2022). For example, the 5-year relative survival  
68 rates of prostate cancer, melanoma, and female breast cancer, during 2011-2017 in the  
69 US were 98%, 93%, and 90%, respectively(Siegel et al., 2022). Such disparity may lead  
70 to an increase in the prevalence and the tumor burden in US society. More importantly,  
71 prolonged survival makes it possible for individuals to be diagnosed with a second  
72 primary cancer (SPC) after the first primary cancer (FPC) during the follow-up.

73 According to the Italian (1976-2010)("Italian cancer figures, report 2013:  
74 Multiple tumours," 2013), the Swiss (1981-2009)(Feller et al., 2020), and the Swedish  
75 (1990-2015)(Zheng et al., 2020) cancer registration data, increased risks of SPCs were  
76 observed in many types of cancer as the FPCs. Patients with oral cavity & pharynx,  
77 larynx, and esophagus as FPC were found to have a significantly elevated risk of any

78 SPCs in both Italy and Switzerland(Feller et al., 2020; "Italian cancer figures, report  
79 2013: Multiple tumours," 2013). In Sweden, liver cancers, as well as nasal and oral  
80 cancers, were found to be associated with SPCs(Zheng et al., 2020). In addition, several  
81 studies focused on certain cancer also suggested a potential relationship between FPC  
82 and SPC(Chattopadhyay et al., 2018). For example, increased FPC risk of colorectal  
83 cancer, kidney cancer, and melanoma were observed following the diagnosis of  
84 non-Hodgkin lymphoma. Despite the strong association observed in these studies,  
85 whether there is any underlying causal relationship is unknown, or the association  
86 observed is due to the potential confounders such as aging.

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

94

## 95 **Results**

96 We set out to identify the observational association between FPCs and SPCs  
97 among 22 solid tumors and 4 hematological malignancies based on the SEER  
98 database (**Figure 1**). A total of 264 significant standardized incidence ratio (SIR)

99 existed among them when compared to a standard population (**Table S1-17**).

100 Hierarchical clustering analysis (heatmap) is shown in **Figure 2** which illustrates the

101 comparison between the cancer incidence in patients with a certain type of FPC and the

102 incidence in the population average level in the SEER dataset. Second primary thyroid

103 cancer, small bowel cancer, or cancers of endocrine system were strongly and

104 significantly associated with most of the FPCs. Cancers of the digestive system,

105 cervix uteri, as well as lung cancer were clustered together. They are closely

106 correlated with an increased risk of several types of cancers such as bladder cancer,

107 kidney cancer, etc. Our subgroup analysis suggested that patients with prostate cancer

108 who received radiation therapy were at an increased risk of being diagnosed with

109 another type of primary cancer during the follow-up including small intestine, soft

110 tissue, and leukemia, compared with those without such therapy (**Table S13**). And

111 SPC risk after breast cancer was inconsistent among males and females. Men with

112 breast cancer had a higher risk for thyroid cancer and prostate cancer ( $SIR_{SEER}=1.30$ ;

113 95% CI, 1.13-1.49;  $SIR_{SEER}=2.33$ , 1.12-4.29; **Table S10**), but with no significant risk

114 change for other cancers.

115 Details of the included exposure-associated SNPs in European ancestry were shown

116 in **Table 1** and **Table S18**. The number of SNPs ranged from 5 (esophagus) to 104

117 (female breast), and the proportion of variance explained by SNPs ( $R^2$ ) ranged from

118 0.10% (thyroid) to 8.46% (bladder). F-statistics for all 13 cancers exceeded 10,

119 suggesting no weak instrument bias here. However, some problems, including too few

120 SNPs, no related GWAS and disease heterogeneity, affected the comprehensive MR  
121 analysis (**Table 1**). Due to the low incidence of cancers in UKB, the power to detect a  
122 significant effective size (0.8/1.2) was relatively low, except for colorectal cancer, lung  
123 cancer, female breast cancer and prostate cancer (**Table S19**).

124 Results from MR analyses are presented in **Table S20-33**. A total of 23 significant  
125 association was detected (16 positive causality and 7 negative causality). Concordant  
126 significant results and unconcordant results between MR and the SEER SIR analyses  
127 are shown in **Table 2** and **Table S20**, respectively. The concordant results suggested  
128 that patients diagnosed with primary oral and pharynx cancer would cause a  
129 significantly increased risk of second primary non-Hodgkin lymphoma ( $SIR_{SEER}=1.18$ ,  
130 95% CI: 1.05-1.31; IVW-MR  $P=8.96\times 10^{-4}$ ). After a primary diagnosis of pancreatic  
131 cancer, SPC risks were increased for small intestine ( $SIR_{SEER}=4.37$ , 95% CI: 2.85-6.40;  
132 MR-Egger  $P=0.04$ ). It also indicated that female patients initially diagnosed with  
133 melanoma would cause a mild but significantly increased risk of cancers in the breast  
134 ( $SIR_{SEER}=1.17$ , 95% CI: 1.12-1.23; IVW-MR  $P=0.04$ ), as well as ovary cancer on soft  
135 tissue related cancer ( $SIR_{SEER}=1.72$ , 95% CI: 1.08-2.60; IVW-MR  $P=8.39\times 10^{-5}$ ). The  
136 greatest number of casual relationships were observed in kidney cancer as FPC. A  
137 primary kidney cancer might cause an elevated risk of cancers of lung and bronchus  
138 ( $SIR_{SEER}=1.28$ , 95% CI: 1.23-1.45; IVW-MR  $P=0.01$ ), non-Hodgkin lymphoma  
139 ( $SIR_{SEER}=1.19$ , 95% CI: 1.08-1.31; IVW-MR  $P=3.64\times 10^{-3}$ ), myeloma ( $SIR_{SEER}=1.54$ ,  
140 95% CI: 1.33-1.78; IVW-MR  $P=3.94\times 10^{-3}$ ). Meanwhile, some primary cancer site

141 might give protective effect against another cancer (for instance, lung cancer vs.  
142 female breast cancer, **Table 2**). Scatter plot, Funnel plot, forest plot and leave-one-out  
143 analysis showed single SNP effective size in **Figure S1-4**, respectively.

144 More conservative analyses were performed to further confirm these causal  
145 relationships. We applied 2 outlier-detected methods with modified second order  
146 weights (radial IVW and radial MR-Egger) and CAUSE (Causal Analysis Using  
147 Summary Effect estimates) to each pair of phenotypes, with the rationale that robust  
148 relationships would exhibit consistent and statistically significant results across  
149 different methods. Additionally, CAUSE is the only method capable of distinguishing  
150 causality from both correlated and uncorrelated pleiotropy. The relationship with at  
151 least 2 significant results was treated as a robust causality (**Table 3**). We found  
152 consistent evidence for a causal effect of oral and pharynx cancer on non-Hodgkin  
153 lymphoma ( $P_{\text{radial-IVW}}=6.00\times 10^{-3}$ ,  $P_{\text{cause}}=8.90\times 10^{-3}$ ), ovary cancer on soft tissue cancer  
154 ( $P_{\text{radial-IVW}}=1.07\times 10^{-3}$ ,  $P_{\text{cause}}=0.01$ ), kidney cancer on lung and myeloma ( $P_{\text{radial-IVW}}=$   
155  $6.60\times 10^{-3}$ ,  $P_{\text{cause}}=0.05$ ;  $P_{\text{radial-IVW}}=0.02$ ,  $P_{\text{cause}}=0.02$ ).

156

## 157 **Discussion**

158 With the expanded life expectancy and the prolonged survival of cancers, the  
159 incidence of SPCs has been rapidly growing in the past decades(Copur & Manapuram,  
160 2019). Genetic factors or shared environmental factors are probably the major causes.  
161 In the previous association studies, individuals with a certain type of primary



162 malignancy would have an increased risk of another malignancy(Feller et al., 2020;  
163 "Italian cancer figures, report 2013: Multiple tumours," 2013; Zheng et al., 2020).  
164 However, whether there are any causal effects within the associations is unclear. In the  
165 present study, via the association study based on the SEER database and the MR  
166 approach using the UKB genetic dataset, we were able to perform this comprehensive  
167 investigation across 26 different types of cancers. 13 out of 26 types of cancers with  
168 adequate GWAS data were able to be further investigated using MR analysis. We  
169 found that numbers of primary malignancies were associated with an increased risk of a  
170 second primary malignancy, however, only a small part of the associations would have  
171 a causal relationship (**Table 2**).

172 Many significant findings were observed in the SEER SIR analysis. SEER is one of  
173 the largest cancer registration-based datasets making itself the most proper data source  
174 to answer the study objectives; however, several advantages of this database and  
175 limitations of the results should be noted. First, the results from the SEER SIR analysis  
176 were associations rather than causal inferences. Many factors may influence the results  
177 of the associations. For example, the confounder of screening effects may exist.  
178 Patients diagnosed with primary cancer might have more frequent healthcare visits  
179 compared to those who did not have any cancers. Therefore, some indolent cancers  
180 such as thyroid cancer and low-risk prostate cancer could have been over-diagnosed  
181 due to the screening effects. Second, it is a cancer registration-based cohort rather than  
182 a population cohort, the standardized incidences calculated from the SEER database

183 may not represent the situation in the general population. Third, treatment preferences  
184 and follow-up strategies may vary in different locations or institutions, which would  
185 also affect the occurrence/detection of the second primary cancer. For example,  
186 radiation therapy may increase the risk of cancer in nearby organs<sup>6</sup>. However, treatment  
187 and follow-up information are not completed in the SEER database due to the natural  
188 design of the cohort<sup>31</sup>. More importantly, lifestyles, comorbidities, and environmental  
189 factors were not included in the SEER database. These are important confounders of  
190 the associations between the first primary cancers and the second primary cancers.

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

202         The lack of GWAS findings would be a major limitation of the MR approach  
203 for some diseases as in the present study. The MR approach may only represent part of

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

220 Finally, besides the limitations mentioned, the relatively small number of cases  
221 of some diseases in UKB may lower our statistical power. As UKB is a  
222 population-based prospective cohort, the relatively short follow-up period may not  
223 allow us to observe enough events (multiple cancers) at this stage. And it is expected to  
224 be independently replicated in another dataset. Regardless, these findings progress our

225 understanding of the relationship underlying both FPCs and SPCs, and potentially  
226 provide points of exploration and intervention that may reduce second primary  
227 cancers.

228

## 229 **Conclusion**

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

234

## 235 **Methods**

### 236 *Study populations*

237 SEER Program 18 Registry database was obtained which covered 27.8% of the  
238 total population in the United States("Number of Persons by Race and Hispanic  
239 Ethnicity for SEER Participants - SEER Registries,."). The SEER program is the  
240 largest cancer incidence dataset in the United States based on population cancer  
241 registration. Based on the ICD-10 code, we identified adult patients (age  $\geq 20$  years)  
242 diagnosed with an FPC between 2000 and 2016, including 22 types of solid-tumor  
243 sites (oral cavity and pharynx, esophagus, stomach, small intestine, colon and rectum,  
244 liver, gallbladder, pancreas, larynx, lung and bronchus, bones and joints, soft tissue  
245 including heart, melanoma of the skin, female breast, male breast, cervix uteri, ovary,

246 prostate, bladder, kidney, renal pelvis and ureter, brain, thyroid) and 4 types of  
247 hematological malignancies (Hodgkin lymphoma, non-Hodgkin lymphoma, myeloma,  
248 leukemia). Patients with diagnosis by autopsy or mentioned in the death certificate only  
249 were excluded.

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

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

262

### 263 *Genotyping and Quality Control*

264 GWAS genotyping array data with imputation and QC from UK Biobank  
265 release V3 was obtained(Bycroft et al., 2018). Briefly, a total of 488,377 participants  
266 (after quality control, QC) were genotyped in UKB using two similar genotyping arrays,

267 the UK Biobank Lung Exome Variant Evaluation (UK BiLEVE with 807,411 markers,  
268 n=49,950) and the Applied Biosystems UK BiLEVE Axiom Array by Affymetrix  
269 (825,927 markers, n=438,427). These two arrays share 95% of the markers. Individuals  
270 were excluded if: (a) ancestry testing using principal component analysis (PCA) to  
271 evaluate the potential conflicts between self-reported ethnicity/race and the genetic  
272 ethnicity/race; (b) Poor call rate at the genotyping stage (n=968, 0.2%); (c) Mismatched  
273 results between self-reported gender and genetic gender (n=652, 0.13%). The genotype  
274 concordance rate was reported as >99.0% (Bycroft et al., 2018). A total of 93,095,623  
275 autosomal SNPs were identified in 487,442 individuals (Bycroft et al., 2018).

276

### 277 *Mendelian Randomization*

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

287 SNPs are independent of each other in terms of linkage disequilibrium (LD,  $r^2 < 0.001$ )  
288 and distance ( $> 10,000$  kb).

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

304 Inverse-variance weighted MR (IVW-MR) and MR-Egger were used in the MR  
305 analyses (Burgess et al., 2019; Burgess & Thompson, 2017; Davies et al., 2018; Hemani  
306 et al., 2018). MR would be performed based on at least 4 SNPs. Briefly, these two  
307 methods are the most used MR methods to infer a causal relationship. IVW-MR is

308 based on a random effect model and is the most efficient with the greatest statistical  
309 power(Hemani et al., 2018). Potential bias as horizontal pleiotropy was evaluated and  
310 adjusted via MR-Egger(Burgess et al., 2019; Burgess & Thompson, 2017; Davies et al.,  
311 2018; Hemani et al., 2018). The causal inference was interpreted via IVW-MR results if  
312 the horizontal pleiotropic effect was not significant; otherwise, based on MR-Egger. A  
313 causal relationship will only be interpreted when a significant MR result was observed  
314 from MR analyses together with a significant association based on the SEER database.  
315 A series of sensitivity and robust analyses including leave-one-out, Radial  
316 MR(Bowden et al., 2019; Bowden et al., 2018) and CAUSE(Morrison et al., 2020)  
317 methods would be performed in case of concordant significant results. MR analyses  
318 were performed using the R package “TwoSampleMR”, “MendelianRandomization”,  
319 “RadialMR” and “cause”.

320

### 321 *Statistical Analysis*

322 The multiple primary standardized incidence ratios (MP-SIR) were defined as the  
323 observed incidence of a second malignancy among cases previously diagnosed with a  
324 certain type of cancer divided by the expected incidence based on the SEER referent  
325 population (the SEER18 2000-2016 referent rate file). All the standardized incidence  
326 observed/expected (O/E) ratios (SIR) and the corresponding 95% confidence intervals  
327 (95% CI) were derived using the “MP-SIR” session of SEER\*Stat (version  
328 8.3.8)(“Multiple Primary - Standardized Incidence Ratios - SEER\*Stat.,”). Only the



329 non-Hispanic white population was included in the present study. We restricted a  
330 minimum 2-month latency period between the first and second primary  
331 diagnosis(Davis et al., 2014). Subgroup analyses were performed after stratifying by  
332 radiation therapy (received or not). All statistical analyses were performed using  
333 SEER\*STAT and R software (4.1.2)(Team, 2021). A type I error of 0.05 (two-sided)  
334 was used to define statistical significance. Multiplicity effects were only considered  
335 during the selection of SNPs.  
336

337 **Reference**

- 338 Bowden, J., Del Greco M, F., Minelli, C., Zhao, Q., Lawlor, D. A., Sheehan, N. A.,  
339 Thompson, J., & Davey Smith, G. (2019). Improving the accuracy of  
340 two-sample summary-data Mendelian randomization: moving beyond the  
341 NOME assumption. *International Journal of Epidemiology*, 48(3), 728-742.  
342 <https://doi.org/10.1093/ije/dyy258>  
343
- 344 Bowden, J., Spiller, W., Del Greco M, F., Sheehan, N., Thompson, J., Minelli, C., &  
345 Davey Smith, G. (2018). Improving the visualization, interpretation and  
346 analysis of two-sample summary data Mendelian randomization via the Radial  
347 plot and Radial regression. *International Journal of Epidemiology*, 47(4),  
348 1264-1278. <https://doi.org/10.1093/ije/dyy101>  
349
- 350 Brion, M.-J. A., Shakhbazov, K., & Visscher, P. M. (2013). Calculating statistical  
351 power in Mendelian randomization studies. *International Journal of*  
352 *Epidemiology*, 42(5), 1497-1501. <https://doi.org/10.1093/ije/dyt179>  
353
- 354 Buniello, A., MacArthur, J. A. L., Cerezo, M., Harris, L. W., Hayhurst, J., Malangone,  
355 C., McMahon, A., Morales, J., Mountjoy, E., Sollis, E., Suveges, D.,  
356 Vrousseau, O., Whetzel, P. L., Amode, R., Guillen, J. A., Riat, H. S.,  
357 Trevanion, S. J., Hall, P., Junkins, H., Flicek, P., Burdett, T., Hindorf, L. A.,  
358 Cunningham, F., & Parkinson, H. (2019). The NHGRI-EBI GWAS Catalog of  
359 published genome-wide association studies, targeted arrays and summary  
360 statistics 2019. *Nucleic Acids Research*, 47(D1), D1005-D1012.  
361 <https://doi.org/10.1093/nar/gky1120>  
362
- 363 Burgess, S., Davey Smith, G., Davies, N. M., Dudbridge, F., Gill, D., Glymour, M.  
364 M., Hartwig, F. P., Holmes, M. V., Minelli, C., Relton, C. L., & Theodoratou,  
365 E. (2019). Guidelines for performing Mendelian randomization investigations.  
366 *Wellcome Open Research*, 4, 186.  
367 <https://doi.org/10.12688/wellcomeopenres.15555.2>  
368
- 369 Burgess, S., & Thompson, S. G. (2017). Interpreting findings from Mendelian  
370 randomization using the MR-Egger method. *European Journal of*  
371 *Epidemiology*, 32(5), 377-389. <https://doi.org/10.1007/s10654-017-0255-x>  
372  
373
- 374 Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L. T., Sharp, K., Motyer, A.,  
375 Vukcevic, D., Delaneau, O., O'Connell, J., Cortes, A., Welsh, S., Young, A.,  
376 Effingham, M., McVean, G., Leslie, S., Allen, N., Donnelly, P., & Marchini, J.

- 377 (2018). The UK Biobank resource with deep phenotyping and genomic data.  
378 *Nature*, 562(7726), 203-209. <https://doi.org/10.1038/s41586-018-0579-z>  
379
- 380 Chattopadhyay, S., Sud, A., Zheng, G., Yu, H., Sundquist, K., Sundquist, J., Försti, A.,  
381 Houlston, R., Hemminki, A., & Hemminki, K. (2018). Second primary  
382 cancers in non-Hodgkin lymphoma: Bidirectional analyses suggesting role for  
383 immune dysfunction. *International Journal of Cancer*, 143(10), 2449-2457.  
384 <https://doi.org/10.1002/ijc.31801>  
385
- 386 Copur, M. S., & Manapuram, S. (2019). Multiple Primary Tumors Over a Lifetime.  
387 *Oncology (Williston Park, N.Y.)*, 33(7).  
388 <https://pubmed.ncbi.nlm.nih.gov/31365752>  
389
- 390 Davies, N. M., Holmes, M. V., & Davey Smith, G. (2018). Reading Mendelian  
391 randomisation studies: a guide, glossary, and checklist for clinicians. *BMJ*  
392 (*Clinical Research ed.*), 362, k601. <https://doi.org/10.1136/bmj.k601>  
393
- 394 Davis, E. J., Beebe-Dimmer, J. L., Yee, C. L., & Cooney, K. A. (2014). Risk of  
395 second primary tumors in men diagnosed with prostate cancer: a  
396 population-based cohort study. *Cancer*, 120(17), 2735-2741.  
397 <https://doi.org/10.1002/cncr.28769>  
398
- 399 Feller, A., Matthes, K. L., Bordoni, A., Bouchardy, C., Bulliard, J.-L., Herrmann, C.,  
400 Konzelmann, I., Maspoli, M., Mousavi, M., Rohrmann, S., Staehelin, K., &  
401 Arndt, V. (2020). The relative risk of second primary cancers in Switzerland: a  
402 population-based retrospective cohort study. *BMC Cancer*, 20(1), 51.  
403 <https://doi.org/10.1186/s12885-019-6452-0>  
404
- 405 Hemani, G., Zheng, J., Elsworth, B., Wade, K. H., Haberland, V., Baird, D., Laurin,  
406 C., Burgess, S., Bowden, J., Langdon, R., Tan, V. Y., Yarmolinsky, J., Shihab,  
407 H. A., Timpson, N. J., Evans, D. M., Relton, C., Martin, R. M., Davey Smith,  
408 G., Gaunt, T. R., & Haycock, P. C. (2018). The MR-Base platform supports  
409 systematic causal inference across the human phenome. *ELife*, 7.  
410 <https://doi.org/10.7554/eLife.34408>  
411
- 412 Hortlund, M., Arroyo Mühr, L. S., Storm, H., Engholm, G., Dillner, J., & Bzhalava, D.  
413 (2017). Cancer risks after solid organ transplantation and after long-term  
414 dialysis. *International Journal of Cancer*, 140(5), 1091-1101.  
415 <https://doi.org/10.1002/ijc.30531>  
416
- 417 Italian cancer figures, report 2013: Multiple tumours. (2013). *Epidemiologia E*  
418 *Prevenzione*, 37(4-5 Suppl 1). <https://pubmed.ncbi.nlm.nih.gov/24259384>

- 419  
420 Kamat, M. A., Blackshaw, J. A., Young, R., Surendran, P., Burgess, S., Danesh, J.,  
421 Butterworth, A. S., & Staley, J. R. (2019). PhenoScanner V2: an expanded  
422 tool for searching human genotype-phenotype associations. *Bioinformatics*  
423 (*Oxford, England*), 35(22), 4851-4853.  
424 <https://doi.org/10.1093/bioinformatics/btz469>  
425
- 426 Morrison, J., Knoblauch, N., Marcus, J. H., Stephens, M., & He, X. (2020).  
427 Mendelian randomization accounting for correlated and uncorrelated  
428 pleiotropic effects using genome-wide summary statistics. *Nature Genetics*,  
429 52(7), 740-747. <https://doi.org/10.1038/s41588-020-0631-4>  
430
- 431 Multiple Primary - Standardized Incidence Ratios - SEER\*Stat.  
432 <https://seer.cancer.gov/seerstat/mp-sir.html>.  
433
- 434 Number of Persons by Race and Hispanic Ethnicity for SEER Participants - SEER  
435 Registries. <https://seer.cancer.gov/registries/data.html#a7>  
436
- 437 Shalpour, S., & Karin, M. (2019). Pas de Deux: Control of Anti-tumor Immunity by  
438 Cancer-Associated Inflammation. *Immunity*, 51(1), 15-26.  
439 <https://doi.org/10.1016/j.immuni.2019.06.021>  
440
- 441 Siegel, R. L., Miller, K. D., Fuchs, H. E., & Jemal, A. (2022). Cancer statistics, 2022.  
442 *CA: a Cancer Journal For Clinicians*, 72(1).  
443 <https://doi.org/10.3322/caac.21708>  
444
- 445 Skrivankova, V. W., Richmond, R. C., Woolf, B. A. R., Yarmolinsky, J., Davies, N.  
446 M., Swanson, S. A., VanderWeele, T. J., Higgins, J. P. T., Timpson, N. J.,  
447 Dimou, N., Langenberg, C., Golub, R. M., Loder, E. W., Gallo, V.,  
448 Tybjaerg-Hansen, A., Davey Smith, G., Egger, M., & Richards, J. B. (2021).  
449 Strengthening the Reporting of Observational Studies in Epidemiology Using  
450 Mendelian Randomization: The STROBE-MR Statement. *JAMA*, 326(16),  
451 1614-1621. <https://doi.org/10.1001/jama.2021.18236>  
452
- 453 Staley, J. R., Blackshaw, J., Kamat, M. A., Ellis, S., Surendran, P., Sun, B. B., Paul, D.  
454 S., Freitag, D., Burgess, S., Danesh, J., Young, R., & Butterworth, A. S.  
455 (2016). PhenoScanner: a database of human genotype-phenotype associations.  
456 *Bioinformatics (Oxford, England)*, 32(20), 3207-3209.  
457 <https://pubmed.ncbi.nlm.nih.gov/27318201>  
458
- 459 Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., &  
460 Bray, F. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of

461 Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: a*  
462 *Cancer Journal For Clinicians*, 71(3), 209-249.  
463 <https://doi.org/10.3322/caac.21660>

464  
465 Team, R. C. (2021). R: A language and environment for statistical computing.  
466 <https://www.R-project.org/>.

467  
468 Zheng, G., Sundquist, K., Sundquist, J., Försti, A., Hemminki, A., & Hemminki, K.  
469 (2020). Rate differences between first and second primary cancers may outline  
470 immune dysfunction as a key risk factor. *Cancer Medicine*, 9(21), 8258-8265.  
471 <https://doi.org/10.1002/cam4.3454>

472

### 473 **Data Availability**

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

477

### 478 **Author Contributions**

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

486

### 487 **Competing interest**

488 The authors declare that there are no competing interests.

489

#### 490 **Acknowledgement**

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

500

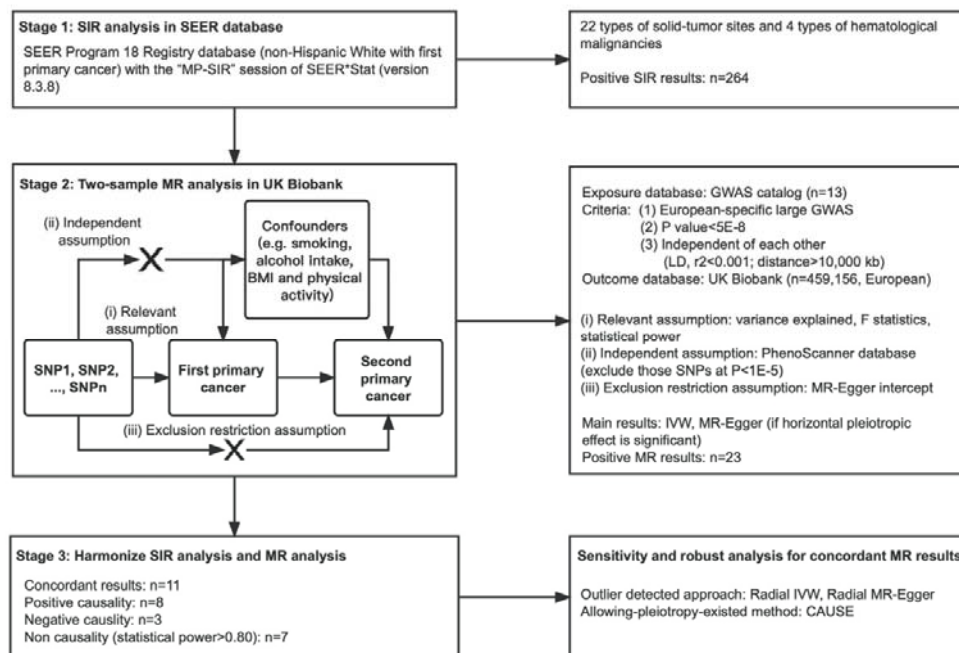
#### 501 **Funding information**

502 This work was in supported by grants from National Natural Science Foundation of  
503 China (Grant No. 81972645), Innovative research team of high-level local universities  
504 in Shanghai, Shanghai Youth Talent Support Program, intramural grant of The  
505 University of Hong Kong to Dr. Rong Na, and Shanghai Sailing Program  
506 (22YF1440500) to Dr. Da Huang. All the funders had no role in study design, data  
507 collection, data analysis, interpretation, and writing of the report.

508

509 **Figure legend**

510 **Figure 1. Study design and work flowchart.**



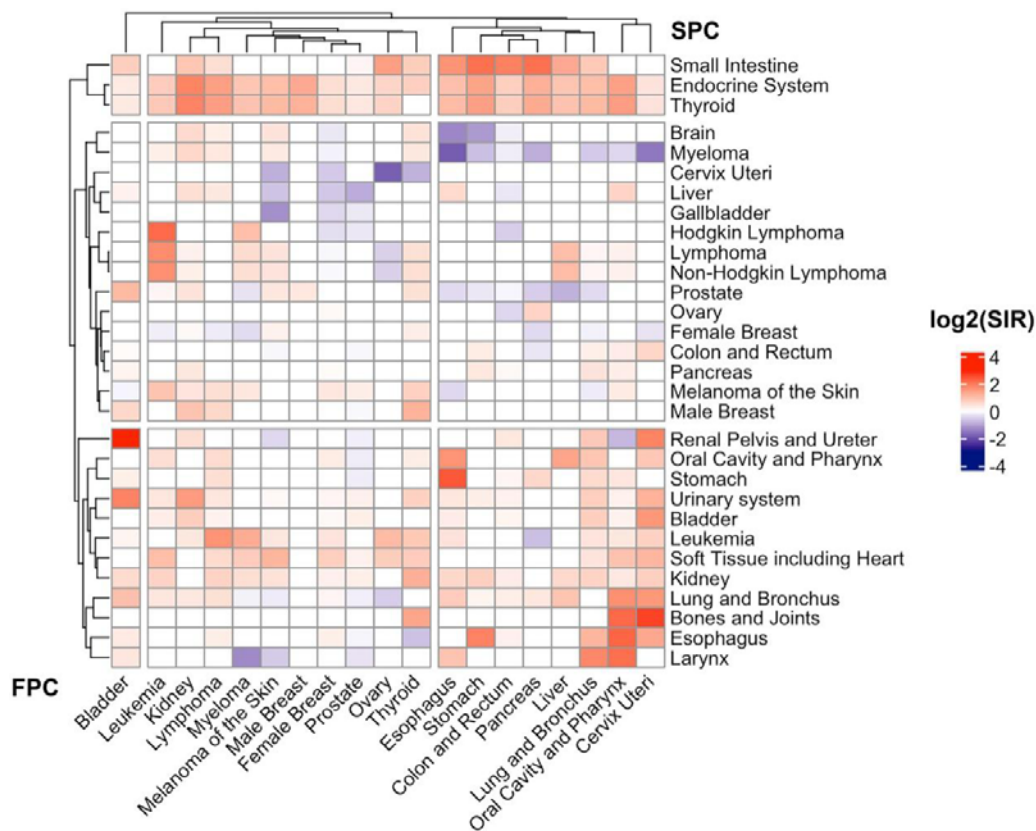
511

512 Abbreviation: SIR, standardized incidence ratio; GWAS, genome-wide significant

513 association.

514

515 **Figure 2. Heat-map of SIR for First primary cancers (FPC, horizontal, cancer**  
 516 **listed in the bottom) and second primary cancers (SPC, vertical, cancer listed on**  
 517 **the right) in the SEER 18 registry (2000-2016).**



518  
 519 The SIR scale is shown in the right corner. Only significant associations (95%  
 520 confidence intervals not overlapping with 1.00) were included; SIRs of the  
 521 insignificant or concordant associations were assigned as 1. Abbreviations: SIR,  
 522 standardized incidence ratios; FPC, first primary cancers; SPC, second primary  
 523 cancers.



**Table 1. Summary of the cancer-specific instrument variables used in this study (European ancestry).**

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

Thyroid	GCST004144	28195142	6	3,001	290,551	0.10	290.83
Myeloma	GCST004483	27363682	13	9,866	249,054	0.31	774.24

---

525 <sup>a</sup> 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  
526 filtering steps discussed in the main text and germline data availability; <sup>b</sup> MR would be performed based on at least 4 SNPs.

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

528 No study found in European: bone and joint, brain, liver, small intestine, gallbladder, renal pelvis and ureter;

529 Heterogeneity: lymphoma, leukemia;

530 Too few SNPs: larynx, stomach, male breast cancer.

531 **Table 2. Concordant causality between Mendelian randomization results and SEER analysis.**

Cancer type (first primary cancer)	Cancer type (second primary cancer)		
	Positive Causality	Negative Causality	No Causality (statistical power $\geq$ 0.80)
Oral and pharynx	Non-hodgkin lymphoma	-	Female breast, Prostate
Esophagus	-	-	-
Colon and rectum	-	-	-
Pancreas	Small intestine	-	Melanoma
Melanoma	Female breast	-	Colon and Rectum
Lung	-	Female Breast	-
Female Breast	-	-	-
Ovary	Soft tissue	-	-
Prostate	-	Colon and Rectum	Non-hodgkin lymphoma
Bladder	-	-	Female breast, Non-hodgkin lymphoma
Kidney	Lung and Bronchus, Melanoma, Non-hodgkin lymphoma, Myeloma	-	-
Thyroid	-	-	-
Myeloma	-	Lung and Bronchus	-

532 See Additional file: Table S20 for unconcordant causality result; Table S21-S33 for the details of Mendelian randomization results (IVW and  
533 MR-Egger).

534 **Table 3. Sensitivity and robust analysis of the concordant causality with outlier-filtering approaches.**

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

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

537