Prognostic evaluation of polygenic risk score underlying pan-cancer analysis: evidence from two large-scale cohorts

Junyi Xin,[a,i] Xia Jiang,[b,c,i] Huiqin Li,[a,d,i] Silu Chen,[a,e] Zhengdong Zhang,[a,e] Meilin Wang,[a,e] Dongying Gu,[f,**] Mulong Du,[d,g,*] and David C. Christiani[g,h,j]

[a]Department of Environmental Genomics, Jiangsu Key Laboratory of Cancer Biomarkers, Prevention and Treatment, Collaborative Innovation Center for Cancer Personalized Medicine, Nanjing Medical University, Nanjing, China
[b]Department of Clinical Neuroscience, Center for Molecular Medicine, Karolinska Institutet, Stockholm, Sweden
[c]Department of Epidemiology and Biostatistics, West China School of Public Health and West China Fourth Hospital, Sichuan University, Chengdu, China
[d]Department of Biostatistics, Center for Global Health, School of Public Health, Nanjing Medical University, Nanjing, China
[e]Department of Genetic Toxicology, The Key Laboratory of Modern Toxicology of Ministry of Education, Center for Global Health, School of Public Health, Nanjing Medical University, Nanjing, China
[f]Department of Oncology, Nanjing First Hospital, Nanjing Medical University, Nanjing, China
[g]Departments of Environmental Health and Epidemiology, Harvard T.H. Chan School of Public Health, 665 Huntington Avenue, Boston, MA, USA
[h]Department of Medicine, Massachusetts General Hospital, 55 Fruit Street, Boston, MA, USA

eBioMedicine
2023;89: 104454
Published Online xxx
https://doi.org/10.1016/j.ebiom.2023.104454

Summary

Background Polygenic risk score (PRS) has been demonstrated to be effective in identifying individuals at high risk of developing cancer, but its prognostic value remains unclear.

Methods We constructed site-specific PRSs by aggregating the risk effect of independent variants derived from previous genome-wide association studies (GWASs) across 17 cancer types. The Cox proportional hazards model was used to evaluate the association of each PRS with cancer survival, leveraging data from two prospective European cohorts, namely the UK Biobank involving 19,628 incident cases and The Cancer Genome Atlas involving 7079 prevalent cases. The combined PRS (CPRS), determined by merging site-specific PRSs, was further used to assess the prognostic effect of PRS on overall cancer in a sex-specific manner.

Findings We discovered that the cancer risk-related PRS was associated with neither overall survival (OS) nor cancer-specific survival (CSS) of each site-specific cancer with an underlying false discovery rate (FDR) $P > 0.05$, as evidenced by consistent findings from the two cohorts. Furthermore, the fixed-effect meta-analysis of the two cohorts provided no evidence to support for an association between CPRS and overall cancer survival in both males [OS: hazard ratio $(HR)_{meta} = 1.00$, $P_{meta} = 0.760$; CSS: $HR_{meta} = 1.01$, $P_{meta} = 0.447$] and females (OS: $HR_{meta} = 0.97$, $P_{meta} = 0.067$; CSS: $HR_{meta} = 0.96$, $P_{meta} = 0.054$). Similar results were observed across multiple sensitivity analyses.

Interpretation Our findings indicate that the risk-specific PRS might not be a clinically useful tool in cancer prognosis prediction and further studies focusing on the development of polygenic prognostic score are warranted.

Funding This project was funded by the National Natural Science Foundation of China (82173601 and 82073631), and Priority Academic Program Development of Jiangsu Higher Education Institutions (Public Health and Preventive Medicine).

Keywords: Polygenic risk score; Cancer; Survival evaluation; Cohort study; Epidemiology

*Corresponding author. Department of Biostatistics, Center for Global Health, School of Public Health, Nanjing Medical University, Nanjing, China.
**Corresponding author.
E-mail addresses: drdumulong@njmu.edu.cn (M. Du), dygu@njmu.edu.cn (D. Gu).
[i]These authors contributed equally to this work.
[j]Senior author who supervised this work.

Research in context

Evidence before this study

Polygenic risk score (PRS) has been demonstrated to be effective in disease risk stratification. Despite it being widely anticipated that PRS can also serve as an indicator of disease outcomes, few studies have assessed its actual prognostic value in clinical practice.

Added value of this study

In this study, we assessed the transferability of PRS in predicting survival across 17 site-specific cancers evidenced by two large-scale longitudinal cohorts, namely, the UK Biobank involving 19,628 incident cases and The Cancer Genome Atlas involving 7079 prevalent cases. We found that the PRS was not associated with either overall survival (OS) or cancer-specific survival (CSS) of site-specific and overall cancer.

Implications of all the available evidence

Results from two large-scale longitudinal cohorts demonstrate a limited clinical utility of risk-based PRS in predicting cancer survival, which emphasizes that a polygenic prognostic score is needed for precision cancer outcome prediction in the future.

Introduction

Cancer is the leading cause of death worldwide, with an estimated 19.3 million new cases and 10.0 million deaths occurring in 2020.[1] It is well known that heritable genetic factors (e.g., genetic variants) and modifiable exposures (e.g., lifestyle) contribute to the development and prognosis of cancer.[2,3] Therefore, a well-developed prediction tool based on risk factors or biomarkers would effectively reduce cancer incidence and mortality.[4]

To date, genome-wide association studies (GWASs) have discovered hundreds of single-nucleotide polymorphisms (SNPs) involved in cancer susceptibility.[5,6] Polygenic risk score (PRS) is a valuable method that aggregates the modest effect of each GWAS-identified SNP and effectively identifies populations at high risk of developing most common cancers, therefore providing more support for the improvement of cancer prevention strategies.[7–9] Of note, it has been suggested that PRS could also serve as a biomarker of disease outcomes, but the actual prognostic value of PRS in clinical practice has not been fully assessed.[10–12] Survival probability is a direct measure of cancer prognosis that is clinically important for patients suffering from cancer,[13] it, however, remains undetermined whether PRS can be utilized to predict cancer survival.

In this study, we performed a pan-cancer survival evaluation of PRS across 17 cancer types (i.e., bladder, brain, breast, colorectal, corpus uteri, esophagus, gastric, lung, lymphoid leukaemia, multiple myeloma, oral and pharynx, ovarian, pancreatic, prostate, renal, skin melanoma and thyroid cancer; Fig. 1), leveraging a total of 26,707 cancer patients of European ancestry with genotyping and clinical information derived from two large-scale longitudinal cohorts, namely the UK Biobank[14] and The Cancer Genome Atlas (TCGA).[15]

Methods

Study subjects

UK biobank

The UK Biobank cohort is a prospective, population-based study that recruited 502,528 adults aged 40–69 years from the general population between April 2006 and December 2010. Participants visited one of 22 assessment centers across England, Scotland and Wales, where they completed touchscreen and nurse-led questionnaires, and provided biological samples. The study protocol and information about data access are available online and more details of the recruitment and study design have been published in previous studies.[14,16] This study was conducted using the UK Biobank Resource under Application #45611.

After individual-level quality control (QC): (i) removed individuals with prevalent cancer (except non-melanoma skin cancer, based on the International Classification of Diseases, 10th revision [ICD-10, C44]) at baseline; (ii) sex discordance; (iii) outliers for genotype missingness or excess heterozygosity; (iv) retained unrelated participants; (v) restricted to "white British" participants of European ancestry and (vi) removed individuals who decided not to participate in this program, a total of 355,543 participants remained for analysis.[17]

The follow-up time of cancer risk was calculated from baseline assessment to the diagnosis of cancer (defined by ICD-10 codes[3]), loss to follow-up, death or end of the follow-up period (December 14, 2016); similarly, the follow-up time of cancer mortality was defined from baseline assessment to death, loss to follow-up or end of the follow-up period. To examine cancer prognosis, the follow-up time of cancer survival was calculated from cancer diagnosis to death or the last follow-up (February 14, 2018), and we determined whether an individual died of a specific cancer by considering the ICD-10 codes listed as the primary cause of death.

TCGA

TCGA is a joint cancer genomics program of the National Cancer Institute and National Human Genome Research Institute that began in 2006. Over the past decade, TCGA collected more than 20,000 primary cancer and matched normal samples from over 10,000 cases across 33 cancer types. TCGA included both tumour and normal biospecimens with written informed consent under

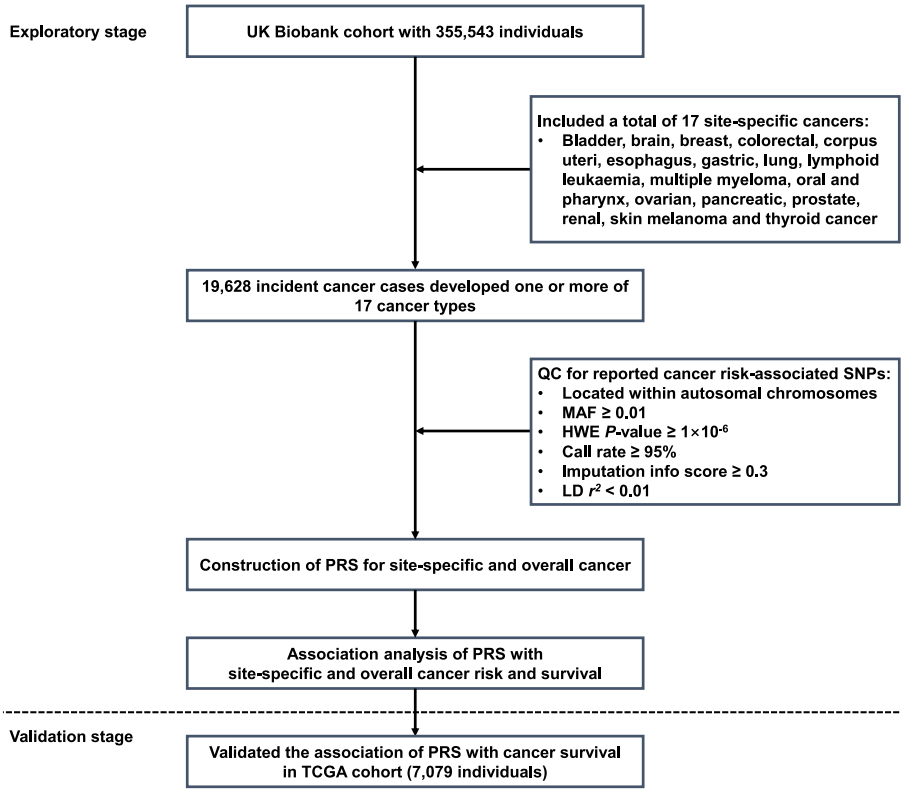

Fig. 1: Summary of the study design. Note: PRS, polygenic risk score; QC, quality control; MAF, minor allele frequency; HWE, Hardy–Weinberg equilibrium; SNP, single nucleotide polymorphism; LD, linkage disequilibrium; TCGA, The Cancer Genome Atlas.

authorization of local institutional review boards. The study protocol and information about data access and study design have been published on the TCGA website (https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga).

After identifying participants of European ancestry,[18] a total of 8523 individuals with 33 cancer types were retained for further analysis. The follow-up time regarding cancer survival was calculated from cancer diagnosis to death or the last follow-up, and detailed clinical information, including overall and cancer-specific death, was extracted from Liu et al.'s study.[19]

Ethics

All participants provided written informed consent prior to data collection. This study was conducted using UK Biobank Resources (Application #45611) and The Cancer Genome Atlas (TCGA). This project was approved by the Internal Review Board of Nanjing Medical University (NJMUIRB-2022-012).

Genotype and imputation

UK biobank

All samples were genotyped using the UK BiLEVE Axiom Array (807,411 markers tested for 49,950 participants) or UK Biobank Axiom Array (825,927 markers tested for 438,427 participants) by Affymetrix. The genotyping data were imputed using SHAPEIT3 and IMPUTE3 based on the reference panels of Haplotype Reference Consortium (HRC), UK10K and 1000 Genomes Project (Phase 3). The study protocol and information about data access are available online (http://www.ukbiobank.ac.uk/wp-content/uploads/2011/11/UK-Biobank-Protocol.pdf).

TCGA

We obtained access to the raw genotype data from normal blood or normal tissue samples in the TCGA database [https://tcga-data.nci.nih.gov/tcga/; the database of Genotypes and Phenotypes (dbGaP) accession phs000178.v11.p8], which included 906,600 SNPs using the Affymetrix SNP 6.0 array. We subsequently imputed the non-genotyped SNPs based on the 1000 Genomes Project (Phase I, version 3, 1092 individuals) using IMPUTE2. The detailed information was reported in our previous study.[20]

Definition of overall survival (OS) and cancer-specific survival (CSS)

OS is defined as the period from the date of diagnosis until the date of death from any cause. The censored

time is from the date of initial diagnosis until the date of last contact (i.e., the largest number of days).

CSS, on the other hand, is defined as the period from the date of initial diagnosis until the date of death caused by the diagnosed cancer. The censored time is from the date of initial diagnosis until the date of death from any cause or until the date of last contact.

Construction of site-specific PRS and sex-specific combined PRS (CPRS)

Based on the following criteria: (i) with genome-wide significant risk loci identified in previous GWASs; and (ii) number of incident cases in UK Biobank >100, a total of 17 cancer types (i.e., bladder, brain, breast, colorectal, corpus uteri, esophagus, gastric, lung, lymphoid leukaemia, multiple myeloma, oral and pharynx, ovarian, pancreatic, prostate, renal, skin melanoma and thyroid cancer) were included in our analysis (Fig. 1). We collected GWAS-reported variants associated with the risk of the 17 cancers from previous studies (Supplementary Table S1a-b)[3] and included variants based on a strict QC process consisting of (i) SNPs located within autosomal chromosomes; (ii) imputation info score ≥0.3; (iii) minor allele frequency (MAF) ≥ 0.01; (iv) call rate ≥95% in the UK Biobank or 90% in the TCGA and (v) Hardy–Weinberg Equilibrium (HWE) P value ≥ 1×10^{-6}.

Subsequently, we constructed PRS for each site-specific cancer based on the remaining independent [linkage disequilibrium (LD) clumping with an $r^2 < 0.01$] SNPs using the following equation: PRS $= \sum_{i=1}^{n} \beta_i SNP_i$, where SNP_i is the risk allele number (0, 1, or 2) of each SNP, and β_i is the logistic regression coefficient derived from previous studies of European ancestry.

Furthermore, to evaluate the association of PRS with overall cancer outcomes, we constructed the sex-specific CPRS using an unweighted method with the following formula: CPRS $= \sum_{c=1}^{C} PRS_c$, where PRS_c is the PRS of the c-th cancer (note: prostate cancer for the male-specific model; breast cancer, ovarian cancer and corpus uteri for the female-specific model).

Sensitivity analysis

To assess the reliability of the association between PRS and cancer survival, we performed several sensitivity analyses: (i) we repeated the evaluation of PRS in 3-year and 5-year cancer survival prediction; (ii) to exclude the joint effects between some risk factors (e.g., smoking) and cancer PRS, we evaluated the association without additional adjustment of body mass index (BMI), smoking status and drinking status in the UK Biobank cohort; (iii) to control for the effect of morbidity severity on cancer survival, we added the site-specific association results with additional adjustment of pathology stage [i.e., American Joint Committee on Cancer (AJCC) stages for most cancer types; clinical stages for corpus uteri and ovarian cancer; and clinical grade for brain cancer] in the TCGA cohort; (iv) to exclude the potential pleiotropy effect of GWAS SNPs between several cancers, we repeated the PRS evaluation after excluding shared variants (LD, $r^2 \geq 0.01$) across all cancers; (v) we further evaluated the prognostic role of genome-wide variants derived PRSs (via a Bayesian approach) of four common cancers (i.e., breast cancer, colorectal cancer, ovarian cancer and prostate cancer; Category: 301; https://biobank.ndph.ox.ac.uk/ukb/label.cgi?id=300)[21] in the UK Biobank cohort; and (vi) to include the epidemiological differences between multiple cancer types, we re-calculated the sex-specific weighted CPRS,[3] with the following formula: CPRS $= \sum_{c=1}^{C} W_c PRS_c$, where W_c is the age-standardized incidence of c-th cancer in UK population, and PRS_c is the PRS of the c-th cancer.

Statistical analysis

The Cox proportional hazards model was used to calculate hazard ratios (HRs) and 95% confidence intervals (CIs) for the associations of each SNP (in the additive genetic model) or PRS with the risk and survival of site-specific cancer, adjusting for relevant confounding factors when appropriate (UK Biobank: sex, age, BMI, smoking status, drinking status and first 10 principal components; TCGA: sex, age and first 10 principal components). We additionally adjusted cancer types when evaluating the association of the CPRS with overall cancer survival. In particular, the logistic regression model was used to calculate odds ratios (ORs) and 95% CIs for the associations of the PRS with site-specific cancer risk in the TCGA cohort, in which the other cancer cases were considered as controls. The heterogeneity between the UK Biobank and TCGA was assessed using Cochrane's Q test, and $P_{\text{heterogeneity}} < 0.01$ was considered significant. The false discovery rate (FDR) method was also performed for multiple comparisons. The Spearman rank correlation analysis was used to measure the relationship of the proportion of deaths or site-specific cancer risk effects between the UK Biobank and TCGA cohorts.

All statistical analyses were performed using R (version 3.6.2) and PLINK (version 1.90) software, and a two-sided P-value less than 0.05 was usually considered significant.

Role of funding source

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

Results

Association of PRS with site-specific cancer survival in the UK biobank

In the UK Biobank cohort, 19,628 of 355,543 individuals were diagnosed with one or more of the 17 included

cancer types during a median follow-up time of 7.8 years, ranging from 179 thyroid cancers to 4882 prostate cancers (Supplementary Table S2); overall, the deaths of 5031 patients were due to all-cause mortality (79.13% attributed to cancer-specific death) during a median follow-up time of 4.1 years after the clinical diagnosis. We collected GWAS-reported SNPs associated with the risk of 17 site-specific cancer types and observed a consistent genetic effect and direction of most SNPs on cancer risk as reported previously. However, no significant association of any individual SNP with OS or CSS of site-specific cancer was observed beyond suggestive genome-wide significance ($P_{Cox} > 1 \times 10^{-6}$; Supplementary Table S1b).

Subsequently, we calculated 17 cancer-specific PRSs by aggregating the risk effect of independent GWAS SNPs and found as expected that all PRSs were significantly associated with an increased risk of site-specific cancer onset (FDR-adjusted $P_{Cox} < 0.05$; Fig. 2a; Supplementary Table S3), with HRs per standard deviation (SD) ranging from 1.14 (gastric cancer) to 2.00 (lymphoid leukaemia). Furthermore, we also observed that most of these PRSs were associated with an increased risk of site-specific cancer death (FDR-adjusted $P_{Cox} < 0.05$; Supplementary Table S4), with the strongest association observed for skin melanoma (HR per SD = 1.89, $P_{Cox} = 5.06 \times 10^{-9}$) and prostate cancer (HR per SD = 1.66, $P_{Cox} = 3.14 \times 10^{-16}$).

However, similar to the finding for individual SNPs, there was no significant association of PRS with either OS or CSS for each site-specific cancer in cancer patients (all FDR-adjusted $P_{Cox} > 0.05$; Fig. 2a; Supplementary Table S3). For example, the colorectal cancer-specific PRS showed significant association with an increased risk of colorectal cancer (HR per SD = 1.56, FDR-adjusted $P_{Cox} = 1.85 \times 10^{-113}$), but not related to the prognosis of colorectal cancer patients (OS: HR per SD = 0.93, FDR-adjusted $P_{Cox} = 0.252$; CSS: HR per SD = 0.96, FDR-adjusted $P_{Cox} = 0.614$).

Association between CPRS and overall cancer survival in the UK biobank

To further evaluate the prognostic effect of the PRS on overall cancer, we integrated site-specific PRSs into sex-specific CPRS due to the differences between males and females in some cancer incidence rates (e.g., prostate cancer). Intriguingly, the CPRS was associated with an increased risk of overall cancer onset in both males (HR per SD = 1.16, $P_{Cox} = 1.02 \times 10^{-52}$) and females (HR per SD = 1.15, $P_{Cox} = 8.72 \times 10^{-40}$; Supplementary Table S3). Moreover, a significant association between CPRS and an increased risk of overall cancer mortality was also observed, especially in males (HR per SD = 1.13, $P_{Cox} = 2.47 \times 10^{-9}$; Supplementary Table S4).

Nevertheless, CPRS was not significantly associated with overall cancer survival in either males (OS: HR per SD = 1.01, $P_{Cox} = 0.599$; CSS: HR per SD = 1.01, $P_{Cox} = 0.532$) or females (OS: HR per SD = 0.99, $P_{Cox} = 0.515$; CSS: HR per SD = 0.97, $P_{Cox} = 0.217$).

Replication of the association of PRS with cancer survival in TCGA

Furthermore, we replicated this evaluation in TCGA cohort that included 7079 cancer patients with 15 cancer types (lymphoid leukaemia and multiple myeloma were not available); with 2366 (33.42%) all-cause deaths and 1678 (23.70%) cancer-specific deaths reported during a median follow-up time of 2.01 years (Supplementary Table S5). The proportion of cancer deaths in the TCGA cohort was positively correlated with that in the UK Biobank cohort ($r_{all\text{-}cause} = 0.671$, $P_{Spearman} = 0.008$; $r_{cancer\text{-}specific} = 0.679$, $P_{Spearman} = 0.007$; Supplementary Fig. S1a and b). Similarly, the PRS was significantly associated with an increased risk of site-specific cancer compared to other cancers [OR per SD > 1; Fig. 2b; Supplementary Table S6]. Notably, there was a high consistency in the association of PRS with site-specific cancer risk between the two cohorts (all $P_{heterogeneity} > 0.01$; Supplementary Fig. S2a), with a high positive correlation of 0.796 ($P_{Spearman} = 6.08 \times 10^{-4}$; Supplementary Fig. S2b). In addition, we observed no significant association between the PRS and site-specific cancer survival in the TCGA cohort (all FDR-adjusted $P_{Cox} > 0.05$; Fig. 2b; Supplementary Table S6), in line with findings from the UK Biobank cohort (all $P_{heterogeneity} > 0.01$; Supplementary Fig. S3a and b).

In terms of CPRS, despite null results with OS and CSS of overall cancer in males (OS: HR per SD = 0.99; $P_{Cox} = 0.789$; CSS: HR per SD = 1.02; $P_{Cox} = 0.664$) or CSS (HR per SD = 0.94; $P_{Cox} = 0.107$) in females, it was marginally significantly associated with an improved OS of overall cancer in females (HR per SD = 0.94; $P_{Cox} = 0.029$; Supplementary Table S6) in the TCGA cohort. However, further fixed-effect meta-analysis combining the association of the CPRS with overall cancer survival in the TCGA and UK Biobank cohorts still yielded to non-significant results in both males (OS: $HR_{meta} = 1.00$, $P_{meta} = 0.760$, $P_{heterogeneity} = 0.614$; CSS: $HR_{meta} = 1.01$, $P_{meta} = 0.447$, $P_{heterogeneity} = 0.960$) and females (OS: $HR_{meta} = 0.97$, $P_{meta} = 0.067$, $P_{heterogeneity} = 0.173$; CSS: $HR_{meta} = 0.96$, $P_{meta} = 0.054$, $P_{heterogeneity} = 0.532$; Fig. 3a and b).

Sensitivity analysis

To assess the reliability of our findings, several sensitivity analyses were performed via different approaches. First, we evaluated the prognostic effect of the PRS on 3-year and 5-year cancer survival. After FDR correction, there was no significant association of the PRS with site-specific cancer survival in the UK Biobank and TCGA cohorts (all FDR-adjusted $P_{Cox} > 0.05$; Supplementary Table S7 and S8), which was consistent with the results from the primary analysis. Similar results were also found between the CPRS and overall

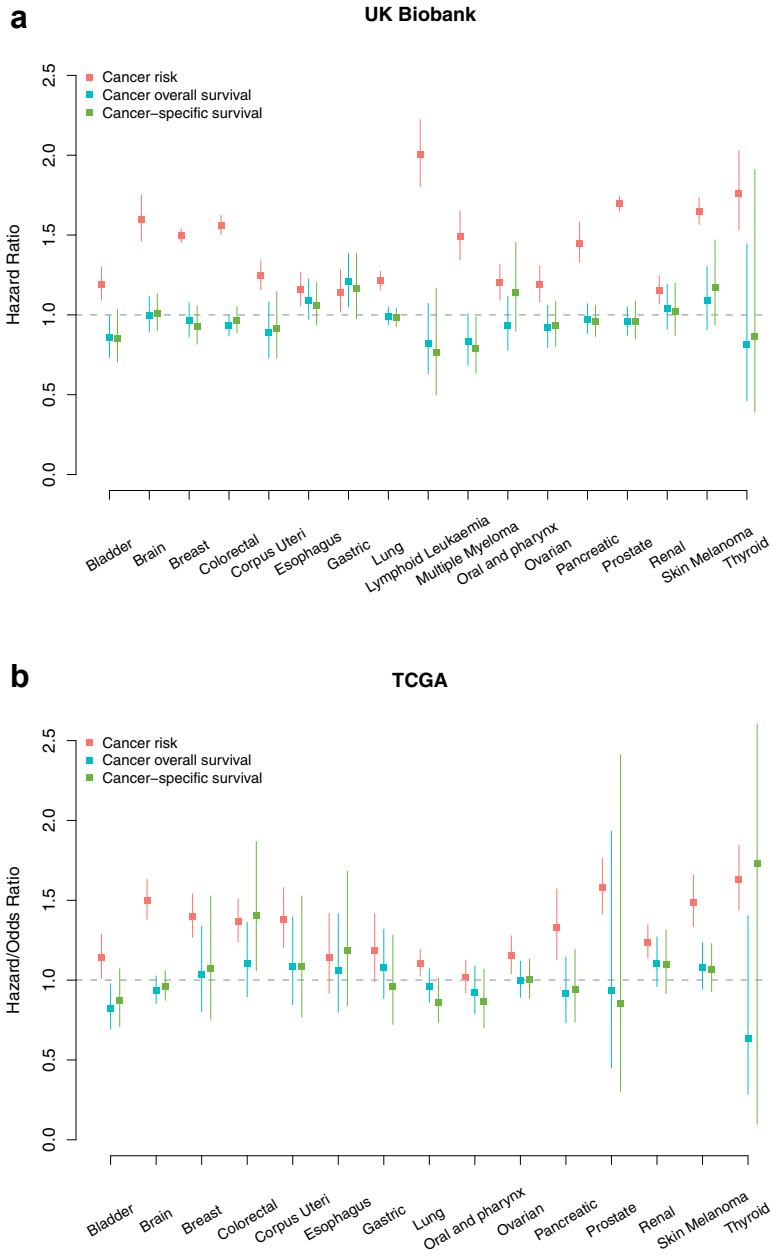

Fig. 2: The associations of PRS with site-specific cancer risk and survival in the UK Biobank and TCGA cohorts. (a) UK Biobank, HR (95% CI) per SD for risk (incident cases/total population) and survival (deaths/cases) evaluation, derived from the Cox proportional hazards regression model with the adjustment of sex, age, BMI, smoking status, drinking status and first 10 principal components when appropriate. (b) TCGA, OR (95% CI) per SD for risk evaluation, derived from the logistic regression model (cancer-specific cases vs. other cancer cases) with the adjustment of sex, age and first 10 principal components when appropriate; HR (95% CI) per SD for survival evaluation, derived from the Cox proportional hazards regression model (deaths/cases) with the adjustment of sex, age and first 10 principal components when appropriate. Note: PRS, polygenic risk score; OR, odds ratio; HR, hazards ratio; CI, confidence interval; TCGA, The Cancer Genome Atlas; SD, standard deviation.

cancer survival (all P_{meta} > 0.05; Supplementary Fig. S4a-d).

Second, when we evaluated the association between PRS and site-specific cancer survival without additional adjustment of BMI, smoking status and drinking status in the UK Biobank cohort, as well as with additional adjustment of pathology stage in the TCGA cohort, similar findings (all FDR-adjusted P_{Cox} > 0.05; Supplementary Tables S9 and S10) were also observed. Third, using independent variants across multiple

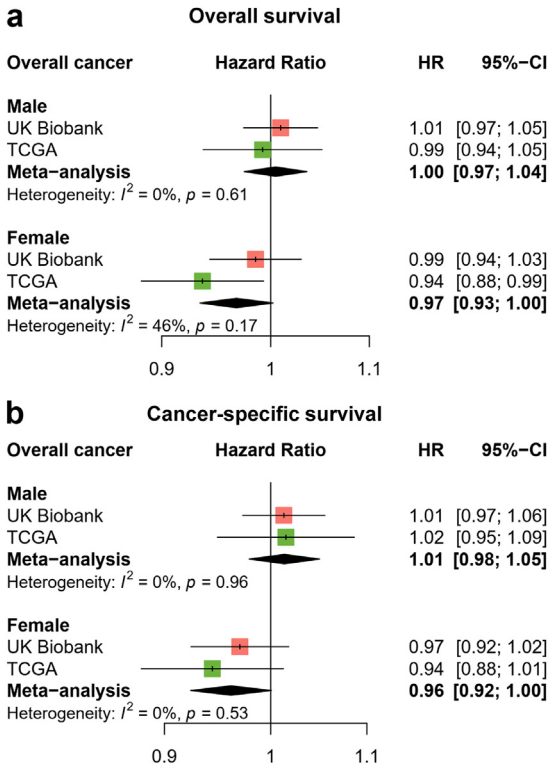

Fig. 3: Meta-analysis of the association of sex-specific CPRS with overall cancer survival between the UK Biobank and TCGA cohorts. (a) Overall survival; (b) cancer-specific survival. HR (95% CI) per SD, derived from the Cox proportional hazards regression model (deaths/cases) with the adjustment of corresponding confounding factors (UK Biobank: age, BMI, smoking status, drinking status, first 10 principal components and cancer types; TCGA: age, first 10 principal components and cancer types) when appropriate. Meta-analysis was performed using a fixed-effect model. Note: CPRS, combined polygenic risk score; HR, hazards ratio; CI, confidence interval; TCGA, The Cancer Genome Atlas; SD, standard deviation.

cancers, we also found non-significant association between PRS and site-specific cancer survival in both cohorts (all FDR-adjusted $P_{Cox} > 0.05$; Supplementary Tables S11 and S12), as well as between the CPRS and overall cancer survival in meta-analysis (all $P_{meta} > 0.05$; Supplementary Fig. S5a and b). Fourth, utilizing PRS derived from genome-wide variants which showed a stronger ability of cancer onset risk prediction, similar null association was found for prognosis (all FDR-adjusted $P_{Cox} > 0.05$; Supplementary Table S13). Lastly, compared to an unweighted CPRS, the weighted CPRS showed a stronger association with an increased risk of overall cancer in both males (HR per SD = 1.29, $P_{Cox} = 7.58 \times 10^{-146}$) and females (HR per SD = 1.25, $P_{Cox} = 1.04 \times 10^{-102}$) in the UK Biobank cohort, yet non-significant associations with prognosis in meta-analysis of two cohorts were still observed among both males and females (all $P_{meta} > 0.05$; Supplementary Fig. S6a and b).

Discussion

In this pan-cancer study, we evaluated the prognostic value of PRS across multiple cancer types, and the consistent results of two large-scale cohorts revealed no significant association of PRS with the prognosis (i.e., OS and CSS) of site-specific and overall cancer.

Currently, PRS has been considered as a robust and cost-efficient tool for translating the findings from GWASs into clinical implications,[22–24] especially for disease risk stratification and has been successfully validated in multiple research studies,[7,25] as well as in our current study. An important advantage of PRS is that compared to conventional risk factors (e.g., lifestyle), PRS is constructed on the basis of inherited genetic variants, which are fixed at conception and can therefore be assessed at an earlier age. It is noteworthy that early genetic risk information may prompt high-risk individuals to make behavioural changes (e.g., exercise, reduced consumption of alcohol and tobacco) to reduce their risk of developing diseases.

Nevertheless, few studies have systematically evaluated the performance of PRS in predicting disease progression and outcomes,[26–28] such as cancer survival. Although cancer survival has been steadily improving in recent decades, it remains a major public health concern.[13] The prediction of cancer survival for individual patients may shed light on individualized cancer therapy that contributes to evaluating tumour behaviour, treatment response, and the patient's ability to withstand tumour burden.[29,30] Here, we performed this comprehensive analysis to determine the prognostic potential of the germline PRS on pan-cancer survival, but we observed no significant association, in agreement with Liu et al.'s findings on the progression of Parkinson's disease[28] and Macauda et al.'s findings on the survival of multiple myeloma.[31] Interestingly, the PRS was significantly associated with an increased risk of cancer-specific mortality among all individuals, demonstrating that PRS was generally more strongly related to cancer onset (e.g., risk and death) than cancer prognosis, which was consistent with a previous study involving multiple diseases.[26] Since most GWASs are designed as case–control studies, rather than by case only for prognosis evaluation, little evidence could support the association between cancer risk-associated SNPs or PRS and cancer survival, as indicated in this study. Importantly, our findings confirm that PRS for risk should not be used for prognosis, in case anyone was considering that. Future GWASs focusing on genetic determinants of cancer progression are thus needed.

This large-scale pan-cancer analysis comprehensively evaluated the relationship of the risk-specific PRS with site-specific and overall cancer survival based on two large-scale prospective cohorts with sufficient sample size. However, several limitations should be acknowledged. First, we only included individuals of European

ancestry for the PRS evaluation, which may miss the ancestral contribution to cancer progression. Second, patients with different cancers usually present diverse survival time, for example, prostate cancer patients usually exhibit longer survival time, while pancreatic cancer patients usually have shorter survival time; therefore, each cancer survival prediction should be evaluated under a sufficient follow-up. Third, cancer survival may also be influenced by the use of individualized therapy (e.g., chemotherapy treatment), and the joint effect between PRS and treatments should be further evaluated. Fourth, we should further evaluate the preferential association of PRS with less aggressive diseases, which may help reveal the potential systematic bias in GWAS and improve the design of future studies.

In summary, we applied PRS to two large-scale cohorts and found evidence that the risk-relevant PRS was not appropriate for cancer survival prediction. The development of further polygenic prognostic score is warranted for cancer progression prediction and precision clinical management.

Contributors

M.D., D.C., D.G., M.W. and J.X. contributed to the conception and design of this study. M.D. and J.X. were involved in the acquisition of data. J.X., X.J., H.L. and S.C. performed analysis. D.C., Z.Z., M.W. and D.G. were involved in data interpretation. All the authors discussed the results and wrote the paper. M.D. and J.X. have verified the underlying data. All authors read and approved the final version of the manuscript.

Data sharing statement

The raw genotype and clinical data have been deposited in UK Biobank (https://www.ukbiobank.ac.uk/) and The Cancer Genome Atlas (TCGA, https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga) program. All relevant data will be shared on reasonable request to the corresponding authors.

Declaration of interests

The authors report no conflict of interest.

Acknowledgements

We are grateful to the participants and study staff of UK Biobank and The Cancer Genome Atlas. This project was funded by the National Natural Science Foundation of China (82173601 and 82073631), and Priority Academic Program Development of Jiangsu Higher Education Institutions (Public Health and Preventive Medicine).

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.ebiom.2023.104454.

References

1 Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2021;71(3):209–249.
2 Schottenfeld D, Beebe-Dimmer JL, Buffler PA, Omenn GS. Current perspective on the global and United States cancer burden attributable to lifestyle and environmental risk factors. *Annu Rev Public Health*. 2013;34:97–117.
3 Zhu M, Wang T, Huang Y, et al. Genetic risk for overall cancer and the benefit of adherence to a healthy lifestyle. *Cancer Res*. 2021;81(17):4618–4627.
4 Soerjomataram I, Bray F. Planning for tomorrow: global cancer incidence and the role of prevention 2020-2070. *Nat Rev Clin Oncol*. 2021;18(10):663–672.
5 Sud A, Kinnersley B, Houlston RS. Genome-wide association studies of cancer: current insights and future perspectives. *Nat Rev Cancer*. 2017;17(11):692–704.
6 Buniello A, MacArthur J, Cerezo M, et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res*. 2019;47(D1):D1005–D1012.
7 Mars N, Koskela JT, Ripatti P, et al. Polygenic and clinical risk scores and their impact on age at onset and prediction of cardiometabolic diseases and common cancers. *Nat Med*. 2020;26(4):549–557.
8 Khera AV, Chaffin M, Aragam KG, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat Genet*. 2018;50(9):1219–1224.
9 Hao L, Kraft P, Berriz GF, et al. Development of a clinical polygenic risk score assay and reporting workflow. *Nat Med*. 2022;28(5):1006–1013.
10 Green ED, Gunter C, Biesecker LG, et al. Strategic vision for improving human health at the Forefront of Genomics. *Nature*. 2020;586(7831):683–692.
11 Lewis CM, Vassos E. Polygenic risk scores: from research tools to clinical instruments. *Genome Med*. 2020;12(1):44.
12 He YQ, Wang TM, Ji M, et al. A polygenic risk score for nasopharyngeal carcinoma shows potential for risk stratification and personalized screening. *Nat Commun*. 2022;13(1):1966.
13 Arnold M, Rutherford MJ, Bardot A, et al. Progress in cancer survival, mortality, and incidence in seven high-income countries 1995-2014 (ICBP SURVMARK-2): a population-based study. *Lancet Oncol*. 2019;20(11):1493–1505.
14 Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature*. 2018;562(7726):203–209.
15 Weinstein JN, Collisson EA, Mills GB, et al. The cancer genome Atlas pan-cancer analysis project. *Nat Genet*. 2013;45(10):1113–1120.
16 Sudlow C, Gallacher J, Allen N, 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.
17 Xin J, Gu D, Chen S, et al. SUMMER: a Mendelian randomization interactive server to systematically evaluate the causal effects of risk factors and circulating biomarkers on pan-cancer survival. *Nucleic Acids Res*. 2022;51(D1):D1160–D1167.
18 Carrot-Zhang J, Chambwe N, Damrauer JS, et al. Comprehensive analysis of genetic ancestry and its molecular correlates in cancer. *Cancer Cell*. 2020;37(5):639–654.
19 Liu J, Lichtenberg T, Hoadley KA, et al. An integrated TCGA pan-cancer clinical data resource to drive high-quality survival outcome analytics. *Cell*. 2018;173(2):400–416.
20 Xin J, Du M, Jiang X, et al. Systematic evaluation of the effects of genetic variants on PIWI-interacting RNA expression across 33 cancer types. *Nucleic Acids Res*. 2021;49(1):90–97.
21 Thompson DJ, Wells D, Selzam S, et al. UK Biobank release and systematic evaluation of optimised polygenic risk scores for 53 diseases and quantitative traits. *medRxiv*. 2022:2022–2026.
22 Torkamani A, Wineinger NE, Topol EJ. The personal and clinical utility of polygenic risk scores. *Nat Rev Genet*. 2018;19(9):581–590.
23 Responsible use of polygenic risk scores in the clinic: potential benefits, risks and gaps. *Nat Med*. 2021;27(11):1876–1884.
24 Lambert SA, Abraham G, Inouye M. Towards clinical utility of polygenic risk scores. *Hum Mol Genet*. 2019;28(R2):R133–R142.
25 Kachuri L, Graff RE, Smith-Byrne K, et al. Pan-cancer analysis demonstrates that integrating polygenic risk scores with modifiable risk factors improves risk prediction. *Nat Commun*. 2020;11(1):6084.
26 Meisner A, Kundu P, Zhang YD, et al. Combined utility of 25 disease and risk factor polygenic risk scores for stratifying risk of all-cause mortality. *Am J Hum Genet*. 2020;107(3):418–431.
27 Landi I, Kaji DA, Cotter L, et al. Prognostic value of polygenic risk scores for adults with psychosis. *Nat Med*. 2021;27(9):1576–1581.
28 Liu G, Peng J, Liao Z, et al. Genome-wide survival study identifies a novel synaptic locus and polygenic score for cognitive progression in Parkinson's disease. *Nat Genet*. 2021;53(6):787–793.
29 Iasonos A, Schrag D, Raj GV, Panageas KS. How to build and interpret a nomogram for cancer prognosis. *J Clin Oncol*. 2008;26(8):1364–1370.
30 Huang S, Yang J, Fong S, Zhao Q. Artificial intelligence in cancer diagnosis and prognosis: opportunities and challenges. *Cancer Lett*. 2020;471:61–71.
31 Macauda A, Clay-Gilmour A, Hielscher T, et al. Does a multiple myeloma polygenic risk score predict overall survival of patients with myeloma? *Cancer Epidemiol Biomarkers Prev*. 2022;31(9):1863–1866.