


Systematic evaluation of cancer-specific genetic risk score for 11 types of cancer in The Cancer Genome Atlas and Electronic Medical Records and Genomics cohorts

Zhuqing Shi^{1,2}  | Hongjie Yu¹ | Yishuo Wu³ | Xiaoling Lin^{2,3} | Quanwa Bao² | Haifei Jia³ | Chelsea Perschon¹ | David Duggan⁴ | Brian T. Helfand¹ | Siqun L. Zheng¹ | Jianfeng Xu^{1,2,3}

¹Program for Personalized Cancer Care, NorthShore University HealthSystem, Evanston, Illinois

²State Key Laboratory of Genetic Engineering, School of Life Science, Fudan University, Shanghai, China

³Fudan Institute of Urology, Huashan Hospital, Fudan University, Shanghai, China

⁴Translational Genomics Research Institute, Phoenix, Arizona

Correspondence

Jianfeng Xu, Program for Personalized Cancer Care, NorthShore University HealthSystem, Evanston, IL
Email: jxu@northshore.org

Funding information

Ellrod-Schweighauser, Chez and Melman families

Abstract

Background: Genetic risk score (GRS) is an odds ratio (OR)-weighted and population-standardized method for measuring cumulative effect of multiple risk-associated single nucleotide polymorphisms (SNPs). We hypothesize that GRS is a valid tool for risk assessment of most common cancers.

Methods: Utilizing genotype and phenotype data from The Cancer Genome Atlas (TCGA) and Electronic Medical Records and Genomics (eMERGE), we tested 11 cancer-specific GRSs (bladder, breast, colorectal, glioma, lung, melanoma, ovarian, pancreatic, prostate, renal, and thyroid cancer) for association with the respective cancer type. Cancer-specific GRSs were calculated, for the first time in these cohorts, based on previously published risk-associated SNPs using the Caucasian subjects in these two cohorts.

Results: Mean cancer-specific GRS in the population controls of eMERGE approximated the expected value of 1.00 (between 0.98 and 1.02) for all 11 types of cancer. Mean cancer-specific GRS was consistently higher in respective cancer patients than controls for all 11 types of cancer ($P < 0.05$). When subjects were categorized into low-, average-, and high-risk groups based on cancer-specific GRS (< 0.5 , 0.5 – 1.5 , and > 1.5 , respectively), significant dose-response associations of higher cancer-specific GRS with higher OR of respective type of cancer were found for nine types of cancer ($P_{\text{trend}} < 0.05$). More than 64% subjects in the population controls of eMERGE can be classified as high risk for at least one type of these cancers.

Conclusion: Validity of GRS for predicting cancer risk is demonstrated for most types of cancer. If confirmed in larger studies, cancer-specific GRS may have the potential for developing personalized cancer screening strategy.

KEYWORDS

age at diagnosis, cancer, genetic risk score

Zhuqing Shi and Hongjie Yu have contributed equally to this work.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2019 The Authors. *Cancer Medicine* published by John Wiley & Sons Ltd.

1 | INTRODUCTION

Cancer is a major public health issue in the United States and across the world. Based on the projection of the National Institute of Health, an estimated 1 735 350 new cases of cancer will be diagnosed in the United States and 609 640 people will die from the disease in 2018.¹ Although most cancer patients do not have germline mutations in known major cancer susceptibility genes, inherited risk factors play an important role in the development of cancer. This notion is supported by many genetic studies, including two large twin studies in Nordic countries.^{2,3} In a prospective study of 80 309 monozygotic and 123 382 same-sex dizygotic twin individuals within the population-based registers of Denmark, Finland, Norway, and Sweden,³ Muccia and colleagues found that heritability (ie, the proportion of variability in disease risk in a population due to genetic factors) of cancer overall was 33%. Significant heritability was observed for the cancer types of skin melanoma (58%), prostate (57%), nonmelanoma skin (43%), ovary (39%), kidney (38%), breast (31%), and corpus uteri (27%). In addition to germline mutations in known cancer susceptibility genes that account for a small proportion of heritability, it is hypothesized that polygenic inheritance (ie, many common but small-effect genetic variants) also contributes significantly to heritability.

Genome-wide association studies (GWAS) in the last decade have successfully identified several hundreds of cancer-specific risk-associated SNPs.^{4,5} Although the biological mechanisms for these SNPs are largely unknown at this stage, the associations are most likely valid due to the stringent criteria for declaring statistical significance ($P < 5 \times 10^{-8}$) and requirement of validation in independent study populations. Individually, these SNPs have a moderate effect on disease risk; with odds ratios (OR) typically ranging from 1.1-1.5. However, when more than one risk-associated SNP is inherited in an individual, they can have a cumulative, clinically significant effect on disease risk.⁶ Polygenic risk scores can now identify a substantially larger fraction of the population at comparable or greater disease risk than is found by rare monogenic mutations.⁷

Several polygenic risk score methods have been employed to measure the cumulative effect of multiple risk-associated SNPs, including (1) a direct risk allele count, (2) an OR-weighted risk allele count, and (3) using the latter approach but with population standardization, commonly termed as a genetic risk score (GRS).⁸ The mean of score from the first two methods will vary depending on the number of risk-associated SNPs used in calculation. In contrast, because GRS is population standardized for each SNP, its expected mean in the general population will always be 1.00 regardless of the number of SNPs used in calculation. Furthermore, GRS values can be simply interpreted as relative risk to the general population. These two

important features of GRS make it easy to implement for individual risk assessment.

Published studies to date have consistently demonstrated associations of various polygenic risk scores with risk for several types of cancer.^{6,9,10} However, associations using the population-standardized GRS have only been reported for a limited number of cancer types such as prostate, breast, and colorectal cancer.^{36,37} We hypothesize that GRS is a valid tool for risk assessment of most common cancers. To test this hypothesis, we systematically assessed associations of 11 cancer-specific GRSs (bladder, breast, colorectal, glioma, lung, melanoma, ovarian, pancreatic, prostate, renal, and thyroid cancer) with their respective cancer risk. This analysis was performed in two large publicly available cohorts: The Cancer Genome Atlas (TCGA) with various types of cancer patients and the Electronic Medical Records and Genomics (eMERGE) Network with a large number of population controls. Results from this study may provide important information for GRS to be used for inherited risk assessment.

2 | METHODS

2.1 | Study subjects and genotyping data

We requested access of these two study cohorts through dbGaP. TCGA is a comprehensive and coordinated effort by the National Institutes of Health (NIH) to accelerate understanding of the molecular basis of cancer through the application of genome analysis technologies, including SNP genotyping. TCGA includes more than 11 000 patients of 33 types of cancer. In this study, we analyzed 11 types of solid tumor cancer where at least six cancer-specific risk-associated SNPs were available. We limited the association analysis in Caucasians due to most study subjects (85%) being of Caucasian descent. Genotyping data from the Affymetrix Genome-Wide Human SNP Array 6.0 are available.

Electronic Medical Records and Genomics is a consortium of five participating sites (Group Health Seattle, Marshfield Clinic, Mayo Clinic, Northwestern University, and Vanderbilt University) funded by the National Health Genome Research Institute (NHGRI) to investigate the use of electronic medical record systems for genomic research.⁴³ The goal of eMERGE is to conduct GWAS in approximately 19 000 individuals using electronic medical record (EMR)-derived phenotypes and DNA from linked biorepositories. Genotyping data from the Illumina Human660W-Quad v1.0 BeadChip are available. Because subjects in eMERGE were not recruited for specific for cancer studies, we treated them as population controls. We did not include a subset of cohort (N = 1700) that was only approved for dementia study. To match race of subjects in TCGA, only Caucasian subjects were included in the analysis (79% of eMERGE subjects were Caucasians).

Cancer type/control group	Sample size (N)	Age at diagnosis (Mean ± SD)	Male (%)
Bladder	343	69 ± 10	74.34%
Breast	827	60 ± 13	0.00%
Colorectal	387	68 ± 13	52.97%
Glioma	992	52 ± 16	58.76%
Lung	908	67 ± 9	60.90%
Melanoma	450	59 ± 16	61.78%
Ovarian	531	60 ± 12	0.00%
Pancreatic	163	66 ± 11	55.21%
Prostate	421	62 ± 7	100.00%
Renal	453	62 ± 12	67.11%
Thyroid	387	49 ± 16	27.39%
eMERGE	13 427	–	47.72%

TABLE 1 Key demographic and clinical information of study subjects

2.2 | Ancestry analysis and SNP imputation

We inferred ancestry information of study subjects in TCGA and eMERGE based on available genotyping data in the SNP arrays using the ADMIXTURE computer program.⁴⁴ Subjects with the estimated proportion of Caucasian ancestry >60% were considered as Caucasians. We also estimated the eigens of these subjects using the EIGENSOFT (Version 3.0) and plotted the first two eigens of these subjects as well as Caucasians, African Americans, and East Asians subjects from the 1000 Genome Project.^{45,46} All Caucasian subjects in the TCGA cohort fell in the cluster of Caucasians (Figure S1).

For risk-associated SNPs that were not included in the downloaded data file, presumably because they were not found on the original genotyping array, imputation was performed using IMPUTE 2.2.2 based on the combined data of the 1000 Genomes Project and HapMap3 data.⁴⁷ A posterior probability of >0.9 was applied to all imputed genotypes.

2.3 | Risk-associated SNPs

Cancer-specific risk-associated SNPs were cataloged based on GWAS papers of the 11 types of cancer published prior to July 1, 2018. The following criteria were used to select independent and reliable risk-associated SNPs: (1) discovered from GWAS studies of Caucasian subjects, with at least 1000 cases and 1000 controls in the first stage; (2) confirmed in additional stages with combined $P < 5 \times 10^{-8}$; and (3) independent, linkage disequilibrium (LD) measurement ($r^2 < 0.2$) between any pair of SNPs. Risk-associated SNPs available directly and indirectly (from imputation) in the TCGA and eMERGE are presented in Table S1, including 10, 66, 30, 19, 6, 17, 11, 9, 79, 10, and 6 SNPs for bladder,^{48,49} breast,^{52,53} colorectal,^{21,59,60} glioma,^{70,71} lung,^{73,74} melanoma,^{78,79}

ovarian,^{84,85} pancreatic,^{89,90} prostate,^{31-33,92} renal,^{97,98} and thyroid cancer,^{102,103} respectively.

2.4 | GRS calculation

GRS, an OR-weighted and population-standardized polygenic risk score, was computed using allelic ORs obtained from the external studies and allele frequencies in the gnomAD (NFE population).⁸ Briefly, GRS was calculated by multiplying the per-allele OR for each SNP and normalized by the expected risk effect of each SNP in the population (W).

$$GRS = \prod_{i=1}^n \frac{OR_i^{g_i}}{W_i}$$

$$W_i = f_i^2 OR_i^2 + 2f_i(1-f_i) OR_i + (1-f_i)^2$$

where, g_i stands for the genotype of SNP i in an individual (0, 1, or 2 risk alleles), OR_i stands for the allelic OR of SNP i , and f_i stands for the risk allele frequency of SNP i . Based on the GRS formula, the mean GRS should be 1.00 in the general population and GRS can be interpreted as relative risk to the general population regardless of the number of SNPs used in the calculation.

2.5 | Statistical analysis

The Wilcoxon rank sum test was used to compare mean cancer-specific GRS in respective cancer patients and controls. Subjects were categorized into low-, average-, and high-risk groups based on their respective cancer-specific GRS (<0.5, 0.5-1.5, and >1.5, respectively). The trend of increasing OR for cancer among subjects in low-, average-,

TABLE 2 Cancer-specific genetic risk score in cases and controls

Cancer type	SNPs (N)	Mean of GRS (95% CI)		P
		Cases	Controls	
Bladder	10	1.04 (1-1.08)	0.98 (0.97-0.98)	3.77E-03
Breast	66	1.15 (1.11-1.2)	1.01 (1-1.03)	1.48E-14
Colorectal	30	1.08 (1.04-1.12)	1 (0.99-1.01)	8.29E-06
Glioma	19	1.22 (1.18-1.26)	0.98 (0.97-0.99)	1.39E-37
Lung	6	1.01 (0.99-1.02)	0.99 (0.98-0.99)	1.16E-02
Melanoma	17	1.2 (1.14-1.26)	1.02 (1.01-1.03)	5.99E-11
Ovarian	11	1.12 (1.08-1.16)	1.02 (1.01-1.03)	1.45E-04
Pancreatic	9	1.13 (1.07-1.18)	1.02 (1.02-1.03)	1.45E-04
Prostate	79	1.3 (1.21-1.38)	0.99 (0.98-1.01)	2.07E-18
Renal	10	1.09 (1.06-1.12)	1.01 (1-1.01)	8.66E-10
Thyroid	6	1.09 (1.04-1.15)	0.98 (0.98-0.99)	3.64E-05

CI, confidence interval; GRS, genetic risk score.

and high-risk groups was tested using a proportion trend test. All statistical tests were performed using R package (Version 3.5.2).

3 | RESULTS

A total of 5871 Caucasian patients diagnosed with one of the 11 types of cancer in the TCGA and 13 427 Caucasian controls from eMERGE were included in this analysis. The key demographic and clinical information for these study subjects are presented in Table 1. For breast and ovarian cancer, only female patients were included and for prostate cancer, only male patients were included.

The mean cancer-specific GRSs approximated the expected value of 1.00 in the population controls of eMERGE for all 11 types of cancer (Table 2); the mean GRSs ranged from 0.98 (glioma bladder, and thyroid cancer) to 1.02 (melanoma, ovarian, and pancreatic cancer). Mean cancer-specific GRS values were significantly higher among respective cancer patients in TCGA than controls in eMERGE for all 11 types of cancer ($P < 0.05$) (Table 2).

Subjects were then categorized into low-, average-, and high-risk groups for each type of cancer based on their respective cancer-specific GRS (<0.5 , $0.5-1.5$, and >1.5 , respectively). Compared to subjects with average-risk, subjects classified as high-risk had OR >1 for their respective type of cancer in 10 types of cancer; nine of which reached statistically significant level ($P < 0.05$) (Table 3). Conversely, compared to subjects with average-risk, subjects classified as low-risk had OR <1 for their respective type of cancer in 10 types of cancer; seven of which reached statistically significant level ($P < 0.05$). A significant dose-response association

of higher cancer-specific GRS with higher odds ratio of respective type of cancer was found for nine types of cancer ($P_{-trend} < 0.05$).

We further estimated the proportion of high-risk subjects in the population controls of the eMERGE cohort. At the individual cancer type level, the proportion of subjects that were classified into high-risk ranged from 2.75% (lung cancer) to 16.15% (prostate cancer) (Table 4). When all 11 types of cancer were tallied together, 64% (61% in male, 66% in female) of subjects were classified as high-risk for at least one type of cancer. 49.50% (49.52% in male, 49.47% in female) of subjects were classified as low-risk for at least one type of cancer, and 84.55% (83.85% in male, 85.19% in female) of subjects were classified as either high-risk or low-risk for at least one type of cancer.

4 | DISCUSSION

This is the first systematic evaluation of cancer-specific and population-standardized GRS for risk assessment of multiple types of cancer and the first study to examine this risk in publicly available study cohorts (TCGA and eMERGE). In a recently published seminal study, Fritche and colleagues studied multiple types of cancer in a large phenome-wide association study (PheWAS) and demonstrated that the top quartiles of cancer-specific polygenic risk score were significantly higher than the bottom quartile for six types of cancer (breast, prostate, melanoma, basal cell carcinoma, squamous cell carcinoma, and thyroid cancer), with OR >2 .⁹ There are many similarities in method, approach, and results between the study described here and their study. Both studies used polygenic risk score methods, adopted multicancer approach, and found evidence that cancer-specific polygenic risk scores

TABLE 3 Odds ratio for each type of cancer among subjects classified as low- and high-risk based on cancer-specific genetic risk score

Cancer type	Low-risk			Average-risk			High-risk		
	Sample size (case/control)	OR (95% CI)	P	Sample size (case/control)	OR	Sample size (case/control)	OR (95% CI)	P	P-trend
Bladder	7/279	1.02 (0.48-2.18)	0.96	301/12245	1.00	35/903	1.58 (1.1-2.25)	0.01	0.02
Breast	68/1064	0.54 (0.42-0.71)	3.42E-06	572/4874	1.00	187/1082	1.47 (1.23-1.76)	1.80E-05	5.02E-15
Colorectal	15/687	0.76 (0.45-1.29)	0.31	324/11324	1.00	48/1416	1.18 (0.87-1.61)	0.28	0.11
Glioma	75/2198	0.48 (0.37-0.61)	9.49E-10	667/9298	1.00	250/1931	1.8 (1.55-2.1)	2.24E-14	4.49E-31
Lung	0/14	0 (0-NaN)	0.33	886/13044	1.00	22/369	0.88 (0.57-1.36)	0.56	0.68
Melanoma	22/1227	0.57 (0.37-0.88)	0.01	323/10262	1.00	105/1938	1.72 (1.37-2.16)	1.80E-06	4.13E-09
Ovarian	10/320	0.43 (0.23-0.82)	0.01	422/5858	1.00	99/842	1.63 (1.3-2.06)	2.64E-05	9.90E-08
Pancreatic	0/399	0 (0-NaN)	0.03	136/11642	1.00	27/1386	1.67 (1.1-2.53)	0.02	9.20E-04
Prostate	36/1274	0.43 (0.3-0.62)	1.76E-06	268/4098	1.00	117/1035	1.73 (1.38-2.17)	1.93E-06	4.02E-16
Renal	1/200	0.15 (0.02-1.08)	0.03	409/12401	1.00	43/826	1.58 (1.14-2.18)	0.01	4.29E-04
Thyroid	21/1063	0.72 (0.46-1.12)	0.15	303/11020	1.00	63/1344	1.7 (1.29-2.25)	1.37E-04	3.55E-05

CI, confidence interval; OR, odds ratio.

are strongly associated with respective cancer risk for multiple types of cancer. However, there is also a major difference in how the two studies actually calculated the polygenic risk score, which can have major implications in interpretation and translation.

Our method uses a population-standardized GRS approach. While this difference—population-standardized versus not—does not affect the performance comparison between cases and controls in a study cohort because the score ranking order of subjects is the same in both methods,⁸ the score values of nonpopulation-standardized methods—for example, top 25%—are not practically meaningful for individuals seen in a clinic. In contrast, because GRS is relative risk to the general population, its values are meaningful for individual subjects and can be used directly to stratify individuals' risk. There are two additional advantages for population-standardized GRS. First, with the expected mean GRS value of 1.00 in the general population, it provides an objective tool to assess the performance of GRS. Deviation from this property signifies a poor performance of GRS. Second, with GRS, the values represent risk compared to the general population, making it straightforward to identify high-risk subjects based on subjects' GRS values.

In this study, we found that the mean cancer-specific GRSs were significantly higher in respective cancer patients than controls for all 11 evaluated types of cancer. When subjects were categorized into low-, average-, and high-risk groups based on their cancer-specific GRSs (<0.5, 0.5-1.5, and >1.5, respectively), a significant dose-response association of higher cancer-specific GRS with higher odds ratio of the respective type of cancer was found for eight types of cancer. Furthermore, we found that the mean GRS values approximated their expected value (1.00) in the population controls of eMERGE for all 11 types of cancer. A significant proportion of subjects (64%) can be classified as high risk (GRS >1.5) for at least one type of cancer in the population controls.

The statistical association of GRS with cancer risk from study populations provides broad-sense validity for its risk stratification. Broad-sense validity is necessary but insufficient to warrant GRS as a testing tool for individual risk assessment. For individual risk assessment, the validity of specific GRS values (we refer to as narrow-sense validity) must be met for several reasons. First, in individual testing, only GRS values of test subjects are available, not the percentiles of GRS that are determined based on all subjects in a study cohort. Clinicians treat patients not cohorts. Second, GRS values, not percentiles, are used directly to estimate an individuals' relative and absolute disease risk including lifetime risk. For example, if a test result provided a prostate cancer GRS value of 1.8 for a 61-year-old Caucasian man, we would report that the subject has a 1.8-fold increased risk for prostate cancer compared to the general population and a 29.6% remaining lifetime risk by age 85 years based

TABLE 4 Proportion of subjects in each risk category in eMERGE

Cancer type	Sample size (N)	Low-risk (GRS <0.5)	Average-risk (GRS:0.5-1.5)	High-risk (GRS >1.5)
Bladder	13 427	2.08%	91.20%	6.73%
Breast	7020	15.16%	69.43%	15.41%
Colorectal	13 427	5.12%	84.34%	10.55%
Glioma	13 427	16.37%	69.25%	14.38%
Lung	13 427	0.10%	97.15%	2.75%
Melanoma	13 427	9.14%	76.43%	14.43%
Ovarian	7020	4.56%	83.45%	11.99%
Pancreatic	13 427	2.97%	86.71%	10.32%
Prostate	6407	19.88%	63.96%	16.15%
Renal	13 427	1.49%	92.36%	6.15%
Thyroid	13 427	7.92%	82.07%	10.01%

GRS, genetic risk score.

on his GRS values, current age, and age-specific incidence and mortality data of Non-Hispanic Whites from SEER data (2011-2015).^{106,107} Therefore, additional evidence related to the narrow-sense validity is needed before GRS can be used in individual risk assessment.

There are important clinical utilities for risk assessment using GRS. For cancer types where a population screening is recommended, such as prostate, breast, colorectal, and lung cancer, primary care physicians can incorporate GRS to develop a personal screening strategy for the need, timing, and frequency of cancer screenings. This personalized approach is likely to maximize the potential benefits and minimize the potential harms of cancer screening.^{109,110} For example, studies from Frampton et al, showed that personalized screening strategy based on polygenic risk score have the potential to greatly reduce the number of individuals screened while still detecting nearly as many cases.^{37,38} For other types of cancer, medical geneticists and specialists can use GRS to supplement other known risk factors, such as family history and high-penetrance genes, to better determine the risk for diagnostic workup.

GRS can be used to supplement family history for a better and more comprehensive assessment of an individuals' risk. These two risk factors have been previously shown to be independent measures of inherited risk. For example, in prostate cancer, family history and a high GRS (>1.4) can identify 17% and 24% of men with high risk for prostate cancer, respectively, in the Prostate Cancer Prevention Trial.⁴⁰ The combination of family history and/or GRS can identify 36% of men at high risk for prostate cancer. The observed prostate cancer risk was 29%, 33%, and 31% for family history alone, GRS alone, and combination of family history and GRS, respectively. GRS has an advantage over family history in that it is an objective measurement of disease risk not susceptible to various issues related to the collection of family history and recall bias. Furthermore, accurate collection of

family history is challenging. For example, family history information of specific cancer was not available in these two important study cohorts (TCGA and eMERGE).

The precise reason for weaker associations of GRS with some types of cancer is unknown but may be due to a number of factors, including fewer numbers of risk-associated SNPs available in this study, and existence of different subtypes of cancer where risk-associated SNPs and etiology could be different. For example, in the lung cancer cohort, 6, 9, and 15 SNPs were reported to be associated with squamous cell, adenocarcinoma, and overall lung cancer, respectively, and some of these SNPs are overlapped. In this study, we calculated lung cancer GRS using risk-associated SNPs reported in any type of lung cancer. This approach was taken because of the limited number of patients available for each subtype of cancer (456 squamous cell lung cancer patients and 452 adenocarcinoma lung cancer patients) and only six risk-associated SNPs in any type of lung cancer were available in both SNP arrays in the TCGA and eMERGE.

A number of additional limitations are noticed in this study. First, the study was limited to Caucasians only, due to the fact that vast majority of study subjects in the TCGA (85%) and eMERGE (79%) are of Caucasian descent. A similar type of analysis should be performed for other racial groups. Second, the sample sizes of patients in TCGA are relatively small, especially for bladder, colorectal, pancreatic, and thyroid cancer (<400). The smaller sample size reduced statistical power in this study. Larger population cohorts and biorepositories, with known case-control status of multiple cancer phenotypes in various racial groups, are needed to replicate and substantiate our findings. For example, data from the PheWAS of Michigan Genomics Initiative can be used to assess GRS performance of multiple types of cancer.⁹ Third, only a subset of established risk-associated SNPs were available in this analysis because genotype data was extracted

from two earlier versions of SNP arrays (Affymetrix Genome-Wide Human SNP Array 6.0 and Illumina Human660W-Quad v1.0 BeadChip). This limitation further reduced the statistical power of our study. Today, low-coverage (~2x) whole-genome sequencing (WGS) is a cost-effective option for obtaining all common variants in the genome, including risk-associated SNPs to be identified in the future.¹¹³

In summary, this study provides additional evidence supporting the use of polygenic risk scores for risk stratification and, specifically, the validity of GRS in predicting cancer risk for several types of cancer. If confirmed in larger studies, cancer-specific GRS may be used for individual risk assessment to develop personalized cancer screening strategy.

ACKNOWLEDGMENTS

We are most grateful to the Ellrodt-Schweighauser, Chez and Melman families for establishing Endowed Chairs of Cancer Genomic Research and Personalized Prostate Cancer Care at NorthShore University HealthSystem in support of Dr. Xu and Dr. Helfand. The results shown here are in whole or part based upon data generated by the TCGA Research Network: <https://www.cancer.gov/tcga>.

CONFLICT OF INTEREST

NorthShore University HealthSystem has an ongoing research agreement with Ambry Genetics to develop GRS for various common diseases.

ORCID

Zhuqing Shi  <https://orcid.org/0000-0001-6321-950X>

REFERENCES

- Noone AMHN, Krapcho M, Miller D, Brest A, Yu M, Ruhl J, Tatalovich Z, Mariotto A, Lewis DR, Chen HS, Feuer EJ, Cronin KA. SEER Cancer Statistics Review, 1975–2015. Bethesda, MD: National Cancer Institute, 2018. https://seer.cancer.gov/csr/1975_2015/.
- Lichtenstein P, Holm NV, Verkasalo PK, et al. Environmental and heritable factors in the causation of cancer—analyses of cohorts of twins from Sweden, Denmark, and Finland. *N Engl J Med*. 2000;343(2):78-85.
- Mucci LA, Hjelmborg JB, Harris JR, et al. Familial risk and heritability of cancer among twins in Nordic countries. *JAMA*. 2016;315(1):68-76.
- Sud A, Kinnnersley B, Houlston RS. Genome-wide association studies of cancer: current insights and future perspectives. *Nat Rev Cancer*. 2017;17(11):692-704.
- MacArthur J, Bowler E, Cerezo M, et al. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res*. 2017;45(D1):D896-D901.
- Zheng SL, Sun J, Wiklund F, et al. Cumulative association of five genetic variants with prostate cancer. *N Engl J Med*. 2008;358(9):910-919.
- 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.
- Conran CA, Na R, Chen H, et al. Population-standardized genetic risk score: the SNP-based method of choice for inherited risk assessment of prostate cancer. *Asian J Androl*. 2016;18(4):520-524.
- Fritsche LG, Gruber SB, Wu Z, et al. Association of polygenic risk scores for multiple cancers in a phenome-wide study: results from the Michigan genomics initiative. *Am J Hum Genet*. 2018;102(6):1048-1061.
- Garcia-Closas M, Rothman N, Figueroa JD, et al. Common genetic polymorphisms modify the effect of smoking on absolute risk of bladder cancer. *Cancer Res*. 2013;73(7):2211-2220.
- Wang P, Ye D, Guo J, et al. Genetic score of multiple risk-associated single nucleotide polymorphisms is a marker for genetic susceptibility to bladder cancer. *Genes Chromosomes Cancer*. 2014;53(1):98-105.
- Reeves GK, Travis RC, Green J, et al. Incidence of breast cancer and its subtypes in relation to individual and multiple low-penetrance genetic susceptibility loci. *JAMA*. 2010;304(4):426-434.
- Sawyer S, Mitchell G, McKinley J, et al. A role for common genomic variants in the assessment of familial breast cancer. *J Clin Oncol*. 2012;30(35):4330-4336.
- Holm J, Humphreys K, Li J, et al. Risk factors and tumor characteristics of interval cancers by mammographic density. *J Clin Oncol*. 2015;33(9):1030-1037.
- Li J, Holm J, Bergh J, et al. Breast cancer genetic risk profile is differentially associated with interval and screen-detected breast cancers. *Ann Oncol*. 2016;27(6):1181.
- Mavaddat N, Pharoah PD, Michailidou K, et al. Prediction of breast cancer risk based on profiling with common genetic variants. *J Natl Cancer Inst*. 2015;107(5):dju036.
- Vachon CM, Pankratz VS, Scott CG, et al. The contributions of breast density and common genetic variation to breast cancer risk. *J Natl Cancer Inst*. 2015;107(5):dju397.
- Li H, Feng B, Miron A, et al. Breast cancer risk prediction using a polygenic risk score in the familial setting: a prospective study from the Breast Cancer Family Registry and kConFab. *Genet Med*. 2017;19(1):30-35.
- Rudolph A, Song M, Brook MN, et al. Joint associations of a polygenic risk score and environmental risk factors for breast cancer in the Breast Cancer Association Consortium. *Int J Epidemiol*. 2018;47(2):526-536.
- Fernandez-Rozadilla C, Kartsonaki C, Woolley C, et al. Telomere length and genetics are independent colorectal tumour risk factors in an evaluation of biomarkers in normal bowel. *Br J Cancer*. 2018;118(5):727-732.
- Schmit SL, Edlund CK, Schumacher FR, et al. Novel common genetic susceptibility loci for colorectal cancer. *J Natl Cancer Inst*. 2019;111(2):146-157.
- Zhao Y, Chen G, Yu H, et al. Development of risk prediction models for glioma based on genome-wide association study findings and comprehensive evaluation of predictive performances. *Oncotarget*. 2018;9(9):8311-8325.

23. Kypreou KP, Stefanaki I, Antonopoulou K, et al. Prediction of melanoma risk in a Southern European population based on a weighted genetic risk score. *J Invest Dermatol*. 2016;136(3):690-695.
24. Cho HG, Ransohoff KJ, Yang L, et al. Melanoma risk prediction using a multilocus genetic risk score in the Women's Health Initiative cohort. *J Am Acad Dermatol*. 2018;79(1):36-41.e10.
25. Cust AE, Drummond M, Kanetsky PA, et al. Assessing the incremental contribution of common genomic variants to melanoma risk prediction in two population-based studies. *J Invest Dermatol*. 2018;138(12):2617-2624.
26. Pearce CI, Rossing MA, Lee AW, et al. Combined and interactive effects of environmental and GWAS-identified risk factors in ovarian cancer. *Cancer Epidemiol Biomarkers Prev*. 2013;22(5):880-890.
27. Li H, Yang L, Zhao X, et al. Prediction of lung cancer risk in a Chinese population using a multifactorial genetic model. *BMC Med Genet*. 2012;13:118.
28. Cheng Y, Jiang T, Zhu M, et al. Risk assessment models for genetic risk predictors of lung cancer using two-stage replication for Asian and European populations. *Oncotarget*. 2017;8(33):53959-53967.
29. Wu Y, Zhang N, Li K, et al. Genetic scores based on risk-associated single nucleotide polymorphisms (SNPs) can reveal inherited risk of renal cell carcinoma. *Oncotarget*. 2016;7(14):18631-18637.
30. Nakatochi M, Lin Y, Ito H, et al. Prediction model for pancreatic cancer risk in the general Japanese population. *PLoS One*. 2018;13(9):e0203386.
31. Al Olama AA, Kote-Jarai Z, Berndt SI, et al. A meta-analysis of 87 040 individuals identifies 23 new susceptibility loci for prostate cancer. *Nat Genet*. 2014;46(10):1103-1109.
32. Schumacher FR, Al Olama AA, Berndt SI, et al. Association analyses of more than 140 000 men identify 63 new prostate cancer susceptibility loci. *Nat Genet*. 2018. 50(7):928-936.
33. Hoffmann TJ, Van Den Eeden SK, Sakoda LC, et al. A large multiethnic genome-wide association study of prostate cancer identifies novel risk variants and substantial ethnic differences. *Cancer Discov*. 2015;5(8):878-891.
34. Grönberg H, Adolfsson J, Aly M, et al. Prostate cancer screening in men aged 50–69 years (STHLM3): a prospective population-based diagnostic study. *Lancet Oncol*. 2015;16(16):1667-1676.
35. Liyanarachchi S, Wojcicka A, Li W, et al. Cumulative risk impact of five genetic variants associated with papillary thyroid carcinoma. *Thyroid*. 2013;23(12):1532-1540.
36. Cuzick J, Brentnall AR, Segal C, et al. Impact of a panel of 88 single nucleotide polymorphisms on the risk of breast cancer in high-risk women: results from two randomized tamoxifen prevention trials. *J Clin Oncol*. 2017;35(7):743-750.
37. Frampton MJ, Law P, Litchfield K, et al. Implications of polygenic risk for personalised colorectal cancer screening. *Ann Oncol*. 2016;27(3):429-434.
38. Frampton M, Houlston RS. Modeling the prevention of colorectal cancer from the combined impact of host and behavioral risk factors. *Genet Med*. 2017;19(3):314-321.
39. Kader AK, Sun J, Reck BH, et al. Potential impact of adding genetic markers to clinical parameters in predicting prostate biopsy outcomes in men following an initial negative biopsy: findings from the REDUCE trial. *Eur Urol*. 2012;62(6):953-961.
40. Chen H, Liu Xu, Brendler CB, et al. Adding genetic risk score to family history identifies twice as many high-risk men for prostate cancer: results from the prostate cancer prevention trial. *Prostate*. 2016;76(12):1120-1129.
41. Ren S, Xu J, Zhou T, et al. Plateau effect of prostate cancer risk-associated SNPs in discriminating prostate biopsy outcomes. *Prostate*. 2013;73(16):1824-1835.
42. Jiang H, Liu F, Wang Z, et al. Prediction of prostate cancer from prostate biopsy in Chinese men using a genetic score derived from 24 prostate cancer risk-associated SNPs. *Prostate*. 2013;73(15):1651-1659.
43. McCarty CA, Chisholm RL, Chute CG, et al. The eMERGE Network: a consortium of biorepositories linked to electronic medical records data for conducting genomic studies. *BMC Med Genomics*. 2011;4:13.
44. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res*. 2009;19(9):1655-1664.
45. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. *PLoS Genet*. 2006;2(12):e190.
46. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*. 2006;38(8):904-909.
47. Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet*. 2007;39(7):906-913.
48. Figueroa JD, Ye Y, Siddiq A, et al. Genome-wide association study identifies multiple loci associated with bladder cancer risk. *Hum Mol Genet*. 2014;23(5):1387-1398.
49. Wu X, Ye Y, Kiemeny LA, et al. Genetic variation in the prostate stem cell antigen gene PSCA confers susceptibility to urinary bladder cancer. *Nat Genet*. 2009;41(9):991-995.
50. Rafnar T, Vermeulen SH, Sulem P, et al. European genome-wide association study identifies SLC14A1 as a new urinary bladder cancer susceptibility gene. *Hum Mol Genet*. 2011;20(21):4268-4281.
51. Rafnar T, Sulem P, Thorleifsson G, et al. Genome-wide association study yields variants at 20p12.2 that associate with urinary bladder cancer. *Hum Mol Genet*. 2014;23(20):5545-5557.
52. Couch FJ, Kuchenbaecker KB, Michailidou K, et al. Identification of four novel susceptibility loci for oestrogen receptor negative breast cancer. *Nat Commun*. 2016;7:11375.
53. Michailidou K, Beesley J, Lindstrom S, et al. Genome-wide association analysis of more than 120 000 individuals identifies 15 new susceptibility loci for breast cancer. *Nat Genet*. 2015;47(4):373-380.
54. Lin WY, Camp NJ, Ghousaini M, et al. Identification and characterization of novel associations in the CASP8/ALS2CR55 region on chromosome 2 with breast cancer risk. *Hum Mol Genet*. 2015;24(1):285-298.
55. Michailidou K, Hall P, Gonzalez-Neira A, et al. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat Genet*. 2013;45(4):353-361, 361e1-2.
56. Gold B, Kirchoff T, Stefanov S, et al. Genome-wide association study provides evidence for a breast cancer risk locus at 6q22.33. *Proc Natl Acad Sci U S A*. 2008;105:4340-4345.
57. Fletcher O, Johnson N, Orr N, et al. Novel breast cancer susceptibility locus at 9q31.2: results of a genome-wide association study. *J Natl Cancer Inst*. 2011;103(5):425-435.

58. Easton DF, Pooley KA, Dunning AM, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature*. 2007;447(7148):1087-1093.
59. Tomlinson IP, Carvajal-Carmona LG, Dobbins SE, et al. Multiple common susceptibility variants near BMP pathway loci GREM1, BMP4, and BMP2 explain part of the missing heritability of colorectal cancer. *PLoS Genet*. 2011;7(6):e1002105.
60. Houlston RS, Cheadle J, Dobbins SE, et al. Meta-analysis of three genome-wide association studies identifies susceptibility loci for colorectal cancer at 1q41, 3q26.2, 12q13.13 and 20q13.33. *Nat Genet*. 2010;42(11):973-977.
61. Peters U, Jiao S, Schumacher FR, et al. Identification of genetic susceptibility loci for colorectal tumors in a genome-wide meta-analysis. *Gastroenterology*. 2013;144(4):799-807.e24.
62. Orlando G, Law PJ, Palin K, et al. Variation at 2q35 (PNKD and TMBIM1) influences colorectal cancer risk and identifies a pleiotropic effect with inflammatory bowel disease. *Hum Mol Genet*. 2016;25(11):2349-2359.
63. Dunlop MG, Dobbins SE, Farrington SM, et al. Common variation near CDKN1A, POLD3 and SHROOM2 influences colorectal cancer risk. *Nat Genet*. 2012;44(7):770-776.
64. Tomlinson IP, Webb E, Carvajal-Carmona L, et al. A genome-wide association study identifies colorectal cancer susceptibility loci on chromosomes 10p14 and 8q23.3. *Nat Genet*. 2008;40(5):623-630.
65. Tomlinson I, Webb E, Carvajal-Carmona L, et al. A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. *Nat Genet*. 2007;39(8):984-988.
66. Tenesa A, Farrington SM, Prendergast J, et al. Genome-wide association scan identifies a colorectal cancer susceptibility locus on 11q23 and replicates risk loci at 8q24 and 18q21. *Nat Genet*. 2008;40(5):631-637.
67. Schumacher FR, Schmit SL, Jiao S, et al. Genome-wide association study of colorectal cancer identifies six new susceptibility loci. *Nat Commun*. 2015;6:7138.
68. Lemire M, Qu C, Loo L, et al. A genome-wide association study for colorectal cancer identifies a risk locus in 14q23.1. *Hum Genet*. 2015;134(11-12):1249-1262.
69. Peters U, Hutter CM, Hsu Li, et al. Meta-analysis of new genome-wide association studies of colorectal cancer risk. *Hum Genet*. 2012;131(2):217-234.
70. Melin BS, Barnholtz-Sloan JS, Wrensch MR, et al. Genome-wide association study of glioma subtypes identifies specific differences in genetic susceptibility to glioblastoma and non-glioblastoma tumors. *Nat Genet*. 2017;49(5):789-794.
71. Shete S, Hosking FJ, Robertson LB, et al. Genome-wide association study identifies five susceptibility loci for glioma. *Nat Genet*. 2009;41(8):899-904.
72. Kinnersley B, Labussière M, Holroyd A, et al. Genome-wide association study identifies multiple susceptibility loci for glioma. *Nat Commun*. 2015;6:8559.
73. Wang Y, McKay JD, Rafnar T, et al. Rare variants of large effect in BRCA2 and CHEK2 affect risk of lung cancer. *Nat Genet*. 2014;46(7):736-741.
74. Fehrer G, Kraft P, Pharoah Pd, et al. Cross-cancer genome-wide analysis of lung, ovary, breast, prostate, and colorectal cancer reveals novel pleiotropic associations. *Cancer Res*. 2016;76(17):5103-5114.
75. McKay JD, Hung RJ, Han Y, et al. Large-scale association analysis identifies new lung cancer susceptibility loci and heterogeneity in genetic susceptibility across histological subtypes. *Nat Genet*. 2017;49(7):1126-1132.
76. Wang Y, Broderick P, Webb E, et al. 5p15.33 and 6p21.33 variants influence lung cancer risk. *Nat Genet*. 2008;40(12):1407-1409.
77. Landi MT, Chatterjee N, Yu K, et al. A genome-wide association study of lung cancer identifies a region of chromosome 5p15 associated with risk for adenocarcinoma. *Am J Hum Genet*. 2009;85(5):679-691.
78. Ransohoff KJ, Wu W, Cho HG, et al. Two-stage genome-wide association study identifies a novel susceptibility locus associated with melanoma. *Oncotarget*. 2017;8(11):17586-17592.
79. Barrett JH, Iles MM, Harland M, et al. Genome-wide association study identifies three new melanoma susceptibility loci. *Nat Genet*. 2011;43(11):1108-1113.
80. Falchi M, Bataille V, Hayward NK, et al. Genome-wide association study identifies variants at 9p21 and 22q13 associated with development of cutaneous nevi. *Nat Genet*. 2009;41(8):915-919.
81. Bishop DT, Demenais F, Iles MM, et al. Genome-wide association study identifies three loci associated with melanoma risk. *Nat Genet*. 2009;41(8):920-925.
82. Nan H, Xu M, Kraft P, et al. Genome-wide association study identifies novel alleles associated with risk of cutaneous basal cell carcinoma and squamous cell carcinoma. *Hum Mol Genet*. 2011;20(18):3718-3724.
83. Brown KM, MacGregor S, Montgomery GW, et al. Common sequence variants on 20q11.22 confer melanoma susceptibility. *Nat Genet*. 2008;40(7):838-840.
84. Pharoah PD, Tsai YY, Ramus SJ, et al. GWAS meta-analysis and replication identifies three new susceptibility loci for ovarian cancer. *Nat Genet*. 2013;45(4):362-370, 370e1-2.
85. Couch FJ, Wang X, McGuffog L, et al. Genome-wide association study in BRCA1 mutation carriers identifies novel loci associated with breast and ovarian cancer risk. *PLoS Genet*. 2013;9(3):e1003212.
86. Goode EL, Chenevix-Trench G, Song H, et al. A genome-wide association study identifies susceptibility loci for ovarian cancer at 2q31 and 8q24. *Nat Genet*. 2010;42(10):874-879.
87. Kuchenbaecker KB, Ramus SJ, Tyrer J, et al. Identification of six new susceptibility loci for invasive epithelial ovarian cancer. *Nat Genet*. 2015;47(2):164-171.
88. Bojesen SE, Pooley KA, Johnatty SE, et al. Multiple independent variants at the TERT locus are associated with telomere length and risks of breast and ovarian cancer. *Nat Genet*. 2013;45(4):371-3384, 384e1-2.
89. Childs EJ, Mocchi E, Campa D, et al. Common variation at 2p13.3, 3q29, 7p13 and 17q25.1 associated with susceptibility to pancreatic cancer. *Nat Genet*. 2015;47(8):911-916.
90. Wolpin BM, Rizzato C, Kraft P, et al. Genome-wide association study identifies multiple susceptibility loci for pancreatic cancer. *Nat Genet*. 2014;46(9):994-1000.
91. Petersen GM, Amundadottir L, Fuchs CS, et al. A genome-wide association study identifies pancreatic cancer susceptibility loci on chromosomes 13q22.1, 1q32.1 and 5p15.33. *Nat Genet*. 2010;42(3):224-228.
92. Kote-Jarai Z, Olama AA, Giles GG, et al. Seven prostate cancer susceptibility loci identified by a multi-stage genome-wide association study. *Nat Genet*. 2011;43(8):785-791.

93. Eeles RA, Olama AA, Benlloch S, et al. Identification of 23 new prostate cancer susceptibility loci using the iCOGS custom genotyping array. *Nat Genet.* 2013;45(4): 385–391, 391e1–2.
94. Al Olama AA, Kote-Jarai Z, Giles GG, et al. Multiple loci on 8q24 associated with prostate cancer susceptibility. *Nat Genet.* 2009;41(10):1058–1060.
95. Gudmundsson J, Sulem P, Gudbjartsson DF, et al. Genome-wide association and replication studies identify four variants associated with prostate cancer susceptibility. *Nat Genet.* 2009;41(10):1122–1126.
96. Yeager M, Orr N, Hayes RB, et al. Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Nat Genet.* 2007;39(5):645–649.
97. Scelo G, Purdue MP, Brown KM, et al. Genome-wide association study identifies multiple risk loci for renal cell carcinoma. *Nat Commun.* 2017;8:15724.
98. Henrion MY, Purdue MP, Scelo G, et al. Common variation at 1q24.1 (ALDH9A1) is a potential risk factor for renal cancer. *PLoS One.* 2015;10(3):e0122589.
99. Purdue MP, Johansson M, Zelenika D, et al. Genome-wide association study of renal cell carcinoma identifies two susceptibility loci on 2p21 and 11q13.3. *Nat Genet.* 2011;43(1):60–65.
100. Henrion M, Frampton M, Scelo G, et al. Common variation at 2q22.3 (ZEB2) influences the risk of renal cancer. *Hum Mol Genet.* 2013;22(4):825–831.
101. Wu X, Scelo G, Purdue MP, et al. A genome-wide association study identifies a novel susceptibility locus for renal cell carcinoma on 12p11.23. *Hum Mol Genet.* 2012;21(2):456–462.
102. Gudmundsson J, Thorleifsson G, Sigurdsson JK, et al. A genome-wide association study yields five novel thyroid cancer risk loci. *Nat Commun.* 2017;8:14517.
103. Gudmundsson J, Sulem P, Gudbjartsson DF, et al. Discovery of common variants associated with low TSH levels and thyroid cancer risk. *Nat Genet.* 2012;44(3):319–322.
104. Gudmundsson J, Sulem P, Gudbjartsson DF, et al. Common variants on 9q22.33 and 14q13.3 predispose to thyroid cancer in European populations. *Nat Genet.* 2009;41(4):460–464.
105. Figlioli G, Kohler A, Chen B, et al. Novel genome-wide association study-based candidate loci for differentiated thyroid cancer risk. *J Clin Endocrinol Metab.* 2014;99(10):E2084–E2092.
106. Fay MP, Pfeiffer R, Cronin KA, Le C, Feuer EJ. Age-conditional probabilities of developing cancer. *Stat Med.* 2003;22(11):1837–1848.
107. Fay MP. Estimating age conditional probability of developing disease from surveillance data. *Popul Health Metr.* 2004;2(1):6.
108. Wen W, Shu XO, Guo X, et al. Prediction of breast cancer risk based on common genetic variants in women of East Asian ancestry. *Breast Cancer Res.* 2016;18(1):124.
109. Grossman DC, Curry SJ, Owens DK, et al. Screening for prostate cancer: US Preventive Services Task Force recommendation statement. *JAMA.* 2018;319(18):1901–1913.
110. Force USPST. Risk assessment, genetic counseling, and genetic testing for BRCA-related cancer in women: recommendation statement. *Am Fam Physician.* 2015;91(2):Online.
111. Moyer VA, Force U. Screening for lung cancer: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med.* 2014;160(5):330–338.
112. Bibbins-Domingo K, Grossman DC, Curry SJ, et al. Screening for colorectal cancer: US Preventive Services Task Force recommendation statement. *JAMA.* 2016;315(23):2564–2575.
113. Rustagi N, Zhou A, Watkins WS, et al. Extremely low-coverage whole genome sequencing in South Asians captures population genomics information. *BMC Genomics.* 2017;18(1):396.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Shi Z, Yu H, Wu Y, et al. Systematic evaluation of cancer-specific genetic risk score for 11 types of cancer in The Cancer Genome Atlas and Electronic Medical Records and Genomics cohorts. *Cancer Med.* 2019;8:3196–3205. <https://doi.org/10.1002/cam4.2143>