Original Investigation | Genetics and Genomics

Genomic Ascertainment of *CHEK2*-Related Cancer Predisposition

Sun Young Kim, MD, PhD; Jung Kim, PhD; Mark Ramos, PhD; Jeremy Haley, MS; Diane Smelser, PhD; H. Shanker Rao, MS; Uyenlinh L. Mirshahi, PhD; Katherine L. Nathanson, MD; Barry I. Graubard, PhD; Hormuzd A. Katki, PhD; David Carey, PhD; Douglas R. Stewart, MD; for the Geisinger-Regeneron DiscovEHR Collaboration and Penn Medicine Biobank

Abstract

IMPORTANCE There is clear evidence that deleterious germline variants in *CHEK2* increase risk for breast and prostate cancers; there is limited or conflicting evidence for other cancers.

OBJECTIVE To quantify the prevalence of as well as cancer risk and survival associated with *CHEK2* germline pathogenic and likely pathogenic variants using genomic ascertainment.

DESIGN, SETTING, AND PARTICIPANTS This case-control study used 2 electronic health record–linked and exome-sequenced biobanks: UK Biobank (n = 469 765) and Geisinger MyCode (adults only; n = 167 050). Variants were classified according to American College of Medical Genetics and Genomics and the Association for Molecular Pathology criteria. Cases were defined as individuals with heterozygous *CHEK2*, harboring pathogenic or likely pathogenic variants; controls as individuals with a benign or likely benign *CHEK2* variation or wildtype *CHEK2*. Cancer registry (MyCode since approximately 1943; UK Biobank since approximately 1970) and demographic data were retrieved; to adjust for relatedness, association analysis was performed with SAIGE-GENE+ with Bonferroni correction.

MAIN OUTCOMES AND MEASURES Prevalence of as well as cancer risk and survival in adults with *CHEK2* germline variants.

RESULTS Of 469 765 individuals in the UK Biobank, there were 3232 case participants (mean [SD] age, 70.8 [8.0] years; 3139 [97.1%] White; 1744 [54.0%] women); of 167 050 individuals with MyCode, there were 3153 case participants (mean [SD] age, 60.5 [17.8] years; 3123 [98.8%] White; 1935 [61.5%] women). In case participants in both MyCode and UKBB, there was a significant excess risk of all cancers (odds ratio [OR], 1.33 [95% CI, 1.18-1.49]; OR, 1.41 [95% CI, 1.26-1.59], respectively), breast (OR, 1.54 [95% CI, 1.18-2.00]; OR, 1.84 [95% CI, 1.49-2.27], respectively), prostate (OR, 1.62 [95% CI, 1.27-2.07]; OR, 1.78 [95% CI, 1.48-2.16], respectively), kidney (OR, 1.58 [95% CI, 1.03-2.41]; OR, 1.84 [95% CI, 1.22-2.77], respectively), and bladder (OR, 1.50 [95% CI, 1.01-2.23]; OR, 1.64 [95% CI, 1.17-2.31], respectively) cancers as well as lymphoid leukemia (OR, 2.08 [95% CI, 1.17-3.69]; OR, 2.21 [95% CI, 1.19-4.08], respectively). Compared with control participants, time to cancer in case participants was significantly shorter in both cohorts; no significant difference was observed between the age-dependent penetrance of truncating and missense variants for cancer in either cohort. Overall survival was significantly decreased in case participants in UK Biobank; however, the primary consequence was seen after 75 years. There was no statistical difference in survival in MyCode. There were no differences in survival between case participants with cancer and control participants with cancer.

CONCLUSIONS AND RELEVANCE In this case-control study of genomic ascertainment of individuals with heterozygous *CHEK2* pathogenic or likely pathogenic variants in 2 population-scale

(continued)

Key Points

Question What is the cancer risk in adults who harbor a pathogenic or likely pathogenic germline *CHEK2* variant when ascertained genomically?

Findings In a case-control study of 2 large population- and health system–based cohorts with a total of 636 815 participants, individuals with heterozygous *CHEK2* variants had a significantly increased risk for all cancer; breast, kidney, bladder, and prostate cancer; and lymphoid leukemia, although cancer risk was generally lower compared with phenotypically ascertained cohorts. There were no significant differences in survival between *CHEK2* case participants with cancer vs control participants with cancer.

Meaning These findings suggest that the method of ascertainment matters when estimating risk and should be considered in clinical decision-making.

+ Supplemental content

Abstract (continued)

cohorts, there was a significant excess risk of breast, prostate, kidney, bladder, and lymphoid leukemia cancer. The conferred excess mortality and cancer risk was low (ORs <2). This has clinical implications for individuals ascertained this way (vs with a family history of cancer).

Introduction

CHEK2 (OMIM 604373) is a tumor-suppressor gene that is involved in DNA repair in response to cellular DNA damage.[1] There is clear evidence that individuals with heterozygous *CHEK2* deleterious germline variants are associated with an increased risk for female breast cancer and prostate cancer, and elevated risks for a variety of other cancers (eg, colorectal, kidney, bladder, leukemia/lymphoma, and thyroid) have been observed.[2,3] In general, germline pathogenic truncating variants (PTV) (eg, c.1100del p.[Thr367fs]) are associated with an increased risk of cancer. In contrast to PTV, pathogenic missense variants (PMV) in *CHEK2* have more variable consequences, mainly dependent on whether a critical protein domain is affected. According to a study by Dorling et al,[4] approximately 60% of rare PMV in *CHEK2* are associated with a lower risk of developing cancers compared with PTV. This suggests that the impact of PMV on cancer susceptibility is not uniform but rather depends on the specific location and nature of the variants. Most work on quantifying risk from a germline variant in a cancer-predisposition gene has arisen from the well-established phenotype-first approach, in which individuals and families are ascertained from their clinical presentation.

Genomic ascertainment is the inversion of the traditional phenotype-first approach.[5] With genomic ascertainment, germline variation of interest is identified, and phenotype status is then obtained from medical records to estimate variant prevalence and disease penetrance and to characterize the phenotype. In principle, this should permit a less biased estimate of the phenotypic spectrum, expressivity, and penetrance of a deleterious variant or set of variants. In this case-control study, we used genomic ascertainment to quantify cancer risk for individuals with heterozygous germline pathogenic or likely pathogenic (P/LP) *CHEK2* variants.

Methods

Setting and Study Participants

In this case-control study, we analyzed 2 population-based cohorts (UK Biobank [UKBB] and Geisinger MyCode) to estimate the prevalence, age-dependent penetrance, cancer risk, and survival of individuals with heterozygous *CHEK2* P/LP variants (case participants) compared with control participants (with non–P/LP *CHEK2* germline variation). Geisinger is an integrated health system serving patients in northeastern and central Pennsylvania, and patients are eligible to participate in the MyCode Community Health Initiative, a system-wide biorepository of blood and DNA samples for broad research purposes.[6] The UKBB is a population-scale biobank.[7]

MyCode participants agree that their samples and data can be linked to Geisinger EHRs; additional informed consent for this study beyond the initial written consent was deemed not to be required per Geisinger Institutional Review board. For the UK Biobank, human participant protection and review was through the North West Multi-Centre Research Ethics Committee. This study followed Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) reporting guideline. Additional details are described in eMethods in Supplement 1.

Sequencing and Relatedness

For UKBB, germline variants were obtained from field 23157, population-level exome OQFE variants, and pVCF format (accessed January 2023). Exome sequencing on UKBB samples has been previously

described.[7,8] The number of unrelated participants was determined by R package ukbtools, using the ukb_gene_samples_to_remove function.

MyCode DNA samples were exome sequenced by the Regeneron Genetics Center as previously described.[9] In the MyCode cohort, we included individuals older than 18 years (n = 167 050). To remove related individuals while maintaining the largest possible cohort, kinship pairs up to third degree relatives (minimum PI_HAT, 0.1875) were used to create a graph of all relatives.

Variant Filtering and *CHEK2* Pathogenicity Classification

All variants that pass quality metrics were annotated using snpEFF,[10] ANNOVAR,[11] ClinVar[12] (database retrieved September 23, 2022), and InterVar version 2.1.3.[13] Variants were classified as P, LP, variant of uncertain significance (VUS), likely benign (LB), and benign (B) using guidelines from the American College of Medical Genetics and Genomics and the Association for Molecular Pathology (ACMG/AMP).[14] Case participants were defined as individuals who harbored a *CHEK2* P/LP variant, whereas control participants included individuals who harbored canonical or B/LB *CHEK2* variation. In this analysis, the all-variant group refers to individuals with all *CHEK2* P/LP variants, PTV refers to those with predicted *CHEK2* truncating P/LP variants, and PMV refers to those with pathogenic missense *CHEK2* P/LP variants. Additional details are in eMethods in Supplement 1.

Cancer Phenotype and Vital Status Query

Demographic data (age, sex, body mass index [BMI], alcohol consumption, smoking history, and self-reported race) were obtained for both case and control participants. All racial groups that were enrolled in the 2 biobanks at the time the data were accessed were included. There were no exclusions on the basis of race or ethnicity. Due to small sample size, groups other than White individuals were collapsed. Clinical phenotypes of neoplasms were obtained from cancer registry for MyCode using *International Statistical Classification of Diseases, Tenth Revision, Clinical Modification* (*ICD-10-CM*) codes. The Cancer Registry (fields 40006 and 40013) and Death Registry data for UKBB (field 40001) were queried using *ICD-9* and *ICD-10* codes.

Statistical Analysis

Demographic data comparisons were completed using Student *t* test for continuous variables and Fisher exact test for binary variables. Power estimates were performed by adapting formulas from Chow et al[15] to a cohort study setting with the assumption of nonbiased ascertainment. Cancer prevalence was modeled using logistic regression with carrier status for all, PTV, and PMV as the main set of explanatory variables and age, sex, smoking history, alcohol consumption, and BMI as covariates. Kaplan-Meier survival analyses were used to estimate all-cause mortality, penetrance of P/LP *CHEK2* variants for cancer, and overall survival for individuals with cancer in MyCode and UKBB cohorts. Hazard ratios were computed using the Cox proportional-hazards (Cox PH) model, adjusting for age, self-reported race, sex, smoking history, alcohol consumption, and BMI, using the log-rank test for equality to compare differences between the curves for control and case groups. Cox PH also adjusted for relatedness by clustering genetically inferred family units. All the analyses were conducted using R version 4.1.2 (R Project for Statistical Computing). Bonferroni correction was applied to organ-system groupings and not specific cancer types; otherwise, *P* < .05 was the level of statistical significance. Additional details on statistical methods are in eMethods in Supplement 1.

Results

Prevalence and Demographic Characteristics of Case Participants With All, PTV, and PMV *CHEK2* Variants in MyCode and UKBB

There were 469 765 individuals in the UKBB cohort and 167 050 individuals in MyCode. eTable 5 in Supplement 2 shows the prevalence of case participants with all variants, PTV, and PMV in both cohorts. eTable 1 in Supplement 2 provides details on the variants. Overall, there were 3232 case

participants (mean [SD] age, 70.8 [8.0] years; 1744 [54.0%] women) in UKBB and 3153 case participants (mean [SD] age, 60.5 [17.8] years; 1935 [61.5%] women) in MyCode (eTable 3 in Supplement 2). Both groups were predominantly White, with 3139 White participants (97.1%) and 93 participants (2.9%) belonging to additional racial groups in UKBB and 3123 White participants (98.8%) and 30 participants (1.0%) belonging to additional racial groups in MyCode. The relatedness (up to the third degree) of the MyCode and UKBB cohorts was approximately 30% and approximately 10%, respectively; eTable 5 in Supplement 2 also shows the heterozygote prevalence in the unrelated fraction of the 2 cohorts. In the 2 cohorts, we observed a 3-fold difference in all *CHEK2* P/LP variant frequency, which is driven by differences in missense variation frequency, particularly I157T (eTable 5 in Supplement 2). Therefore, we investigated the frequency of 6 common *CHEK2* variants in gnomAD version 4.1, Penn Medicine Biobank (PMBB), and All of Us.[16] The 2 US-based biobanks (PMBB and All of Us[16]) had similar frequencies for those variants with Geisinger (eTable 2 in Supplement 2). Similarly, the Regeneron Million Exome Variant Browser version 1.1.3[17,18] showed a greater than 10-fold difference in allele frequency for *CHEK2* I157 in populations in Northern Europe vs the British Isles. eTable 3 in Supplement 2 lists demographic and covariate data for case participants with all variants, PTV, and PMV as well as control participants.

Risk for Cancer in Both MyCode and UKBB

Figure 1A displays statistically significant associations for case participants with all variants, PTV, and PMV in *CHEK2* with organ system groupings of cancer in MyCode. The odds ratios (ORs) and Bonferroni-corrected *P* values for case participants are shown. In case participants in the all-variant group, there was a significant excess risk of all cancers (OR, 1.33 [95% CI, 1.18-1.49]), breast cancer (*ICD-10-CM* code C50: OR, 1.54 [95% CI, 1.18-2.00]), male genital organ cancer (*ICD-10-CM* codes C60-C63), urinary tract cancer (*ICD-10-CM* codes C64-C68), thyroid and other endocrine gland cancers (*ICD-10-CM* codes C73-C75), and lymphoid, hematopoietic, and related tissue cancer (*ICD-10-CM* codes C81-C96). (Of all C50 codes observed in case participants, 155 of 156 [99.4%] and 225 of 227 [99.1%] were in female individuals and 1 [0.6%] and 2 [0.8%] were in males in MyCode and UKBB, respectively.) eFigure 1 in Supplement 1 displays the ORs for the MyCode cohort for all-variant, PTV, and PMV case participant groups for all organ system groupings of cancer *ICD* codes. Figure 1B displays the ORs for the UKBB cohort for all-variant, PTV, and PMV groups of case participants for organ system groupings of cancer *ICD* codes with a significant excess of risk. In the all groups of case participants, there was a significant excess risk of developing all cancers (OR, 1.41 [95% CI, 1.26-1.59]), breast cancer (*ICD-10-CM* code C50; OR, 1.84 [95% CI, 1.49-2.27]), male genital organ cancer (*ICD-10-CM* codes C60-C63), urinary tract cancer (*ICD-10-CM* codes C64-C68), cancer from the secondary and unspecified sites (*ICD-10-CM* codes C76-C79), and lymphoid, hematopoietic, and related tissue cancer (*ICD-10-CM* codes C81-C96). In contrast to MyCode, there was no statistically significant excess risk to develop thyroid and other endocrine gland cancer (*ICD-10-CM* codes C73-C75). eFigure 2 in Supplement 1 displays the ORs for the UKBB cohort for the all, PTV, and PMV case participant groups for all organ system groupings of cancer *ICD* codes.

Specific Cancers Associated With Case Participants With All Variants, PTV, and PMV in *CHEK2*

Figure 2A shows the specific types of cancer in the MyCode cohort with an excess risk from the significant organ-system analysis shown in Figure 1A. Of note is the significant excess risk for prostate cancer (C61: OR, 1.62 [95% CI, 1.27-2.07]), kidney cancer (C64: OR, 1.58 [95% CI, 1.03-2.41]), bladder cancer (C67: OR, 1.50 [95% CI, 1.01-2.23]), thyroid cancer (C73: OR, 2.04 [95% CI, 1.48-2.82]), and lymphoid leukemia (C91: OR, 2.08 [95% CI, 1.17-3.69]) in the all-variant group of case participants. eFigure 3 in Supplement 1 displays the ORs for the MyCode cohort for all-variant, PTV, and PMV *CHEK2* groups for all specific types of cancer from all organ system groupings of cancer *ICD* codes. eTable 4 in Supplement 2 lists the case counts and percentages for PMV, PTV, and all-variant groups in the case cohort and fold-enrichment (vs controls) for each of the *ICD-10* diagnostic codes in MyCode.

Figure 1. Cancer Risk by Organ System for Case Participants in Both Cohorts

A MyCode

ICD-10-CM code and cancer type	Participants, No. %	OR (95% CI)	Adjusted SAIGE P value
All cancers			
Control	34 160 (22.40)	1 [Reference]	NA
Case			
All	864 (27.40)	1.33 (1.18-1.49)	2.15×10^{-10}
PMV	609 (27.40)	1.38 (1.11-1.71)	2.84×10^{-7}
PTV	247 (27.10)	1.30 (1.12-1.49)	1.09×10^{-4}
C50 Malignant neoplasm of breast			
Control	5004 (3.28)	1 [Reference]	NA
Case			
All	156 (4.95)	1.54 (1.18-2.00)	2.45×10^{-5}
PMV	105 (4.73)	1.45 (1.07-1.97)	1.05×10^{-2}
PTV	49 (5.37)	1.73 (1.10-2.73)	.01
C60-C63 Malignant neoplasms of male genital organs			
Control	4443 (7.49)	1 [Reference]	NA
Case			
All	142 (11.66)	1.61 (1.21-2.15)	2.05×10^{-4}
PMV	106 (12.08)	1.68 (1.22-2.31)	6.90×10^{-4}
PTV	36 (10.74)	1.46 (0.84-2.54)	.96
C64-C68 Malignant neoplasms of urinary tract			
Control	2632 (1.72)	1 [Reference]	NA
Case			
All	84 (2.66)	1.56 (1.11-2.20)	3.76×10^{-3}
PMV	53 (2.39)	1.35 (0.89-2.05)	.70
PTV	31 (3.40)	2.16 (1.23-3.78)	.03
C73-C75 Malignant neoplasms of thyroid and other endocrine glands			
Control	1840 (1.21)	1 [Reference]	NA
Case			
All	66 (2.09)	1.76 (1.20-2.56)	9.11×10^{-4}
PMV	49 (2.21)	1.86 (1.21-2.85)	1.11×10^{-3}
PTV	16 (1.75)	1.45 (0.69-3.06)	>.99
C81-C96 Malignant neoplasms, stated or presumed to be primary, of lymphoid, hematopoietic, and related tissue			
Control	3107 (2.04)	1 [Reference]	NA
Case			
All	91 (2.88)	1.42 (1.03-1.95)	.04
PMV	60 (2.70)	1.30 (0.88-1.92)	>.99
PTV	31 (3.40)	1.74 (1.01-3.01)	.06

OR (95% CI) axis: 0.3　0.5　1　4.0

B UK Biobank

ICD-10-CM code and cancer type	Participants, No. (%)	OR (95% CI)	Adjusted SAIGE P value
All cancers			
Control	69 836 (22.87)	1 [Reference]	NA
Case			
All	934 (28.90)	1.41 (1.26-1.59)	6.05×10^{-15}
PMV	344 (26.67)	1.28 (1.06-1.55)	2.19×10^{-4}
PTV	562 (30.23)	1.50 (1.28-1.75)	9.56×10^{-12}
C50 Malignant neoplasm of breast			
Control	12 439 (4.07)	1 [Reference]	NA
Case			
All	227 (7.02)	1.84 (1.49-2.27)	2.47×10^{-12}
PMV	80 (6.20)	1.55 (1.09-2.19)	1.41×10^{-2}
PTV	139 (7.53)	2.05 (1.57-2.69)	4.49×10^{-10}
C60-C63 Malignant neoplasms of male genital organs			
Control	10 557 (3.46)	1 [Reference]	NA
Case			
All	179 (5.53)	1.77 (1.39-2.25)	3.64×10^{-9}
PMV	58 (4.50)	1.52 (1.01-2.30)	.14
PTV	113 (6.12)	1.87 (1.38-2.53)	4.13×10^{-7}
C64-C68 Malignant neoplasms of urinary tract			
Control	4351 (1.43)	1 [Reference]	NA
Case			
All	77 (2.38)	1.75 (1.24-2.46)	1.01×10^{-4}
PMV	27 (2.09)	1.58 (0.89-2.80)	.84
PTV	49 (2.65)	1.92 (1.25-2.94)	2.68×10^{-4}
C76-C79 Malignant neoplasms of ill-defined, secondary and unspecified sites			
Control	14 339 (4.70)	1 [Reference]	NA
Case			
All	192 (5.94)	1.31 (1.05-1.63)	4.68×10^{-3}
PMV	66 (5.12)	1.13 (0.78-1.64)	>.99
PTV	120 (6.50)	1.43 (1.09-1.90)	2.95×10^{-3}
C81-C96 Malignant neoplasms, stated or presumed to be primary, of lymphoid, hematopoietic, and related tissue			
Control	5030 (1.65)	1 [Reference]	NA
Case			
All	108 (3.34)	2.11 (1.58-2.82)	3.45×10^{-9}
PMV	39 (3.02)	1.95 (1.21-3.15)	4.70×10^{-3}
PTV	65 (3.52)	2.19 (1.51-3.17)	8.74×10^{-7}

OR (95% CI) axis: 0.3　0.5　1　4.0

Odds ratio (ORs) for case participants with any pathogenic or likely pathogenic variant (all), those with truncating pathogenic or likely pathogenic variants (PTV), and those with pathogenic missense variants (PMV) for organ system groupings of cancer *International Classification of Diseases* codes with a significant excess of risk in MyCode (A) and UK Biobank (B). NA indicates not applicable; SAIGE, Scalable and accurate implementation of generalized mixed model.

Figure 3 shows the specific types of cancer in the UKBB cohort with an excess risk from the organ-system analysis shown in Figure 1B. Of note is the significant excess risk for prostate (C61, all-variant group: OR, 1.78 [95% CI, 1.48-2.16]), kidney cancer (C64, all-variant group: OR, 1.84 [95% CI, 1.22-2.77]), and bladder cancer (C67, all-variant group: OR, 1.64 [95% CI, 1.17-2.31]) in all-variant and PTV groups of case participants. There was significant increased risk for diffuse non-Hodgkin lymphoma (C83: OR, 1.84 [95% CI, 1.06-3.20]), other and nonspecified types of non-Hodgkin lymphoma (C85: OR, 1.83 [95% CI, 1.08-3.08]) and lymphoid leukemia (C91: OR, 2.21 [95% CI, 1.19-4.08]) in the all-variant group, whereas peripheral and cutaneous T-cell lymphomas (C84) were exclusively associated with case participants with PMV. eFigure 4 in Supplement 1 displays the ORs for the UKBB cohort for the all-variant, PTV, and PMV *CHEK2* groups for all specific types of cancer from all organ system groupings of cancer *ICD* codes. eTable 4 in Supplement 2 lists the case counts and percentages for PMV, PTV, and all-variant groups for the case cohort and fold-enrichment (vs controls) for each of the *ICD-10* diagnostic codes in UKBB.

Time to Cancer

Compared with controls, time to all cancer in the all group of case participants was significantly different in both MyCode (adjusted HR, 1.26 [95% CI, 1.17-1.36]; *P* < .001) and UKBB (adjusted HR, 1.31 [95% CI, 1.24-1.40]; *P* < .001) (**Figure 4**A and **Figure 5**A). Case participants in the PMV group were at higher risk for all cancers tested compared with control participants in both MyCode

Figure 2. Organ System–Specific Cancer Risks for Case Participants in MyCode

| *ICD-10-CM* code and cancer type	Participants, No. (%)	OR (95% CI)		Adjusted SAIGE *P* value
C61 Malignant neoplasm of prostate				
Control	4222 (7.12)	1 [Reference]		NA
Case				
All	135 (11.09)	1.62 (1.27-2.07)		1.04×10^{-4}
PMV	101 (11.51)	1.69 (1.27-2.24)		4.03×10^{-4}
PTV	34 (10.14)	NA		NA
C64 Malignant neoplasm of kidney, except renal pelvis				
Control	1202 (0.79)	1 [Reference]		NA
Case				
All	39 (1.24)	1.58 (1.03-2.41)		.02
PMV	27 (1.22)	NA		NA
PTV	12 (1.31)	1.76 (0.83-3.75)		.28
C67 Malignant neoplasm of bladder				
Control	1477 (0.96)	1 [Reference]		NA
Case				
All	46 (1.46)	1.50 (1.01-2.23)		.10
PMV	27 (1.22)	NA		NA
PTV	19 (2.08)	2.34 (1.25-4.38)		3.51×10^{-3}
C73 Malignant neoplasm of thyroid gland				
Control	1370 (0.90)	1 [Reference]		NA
Case				
All	57 (1.81)	2.04 (1.48-2.82)		1.82×10^{-5}
PMV	42 (1.89)	2.15 (1.47-3.14)		6.60×10^{-5}
PTV	14 (1.53)	NA		NA
C91 Lymphoid leukemia				
Control	627 (0.41)	1 [Reference]		NA
Case				
All	27 (0.86)	2.08 (1.17-3.69)		7.61×10^{-3}
PMV	19 (0.86)	NA		NA
PTV	8 (0.88)	2.21 (0.78-6.22)		.53

OR (95% CI) axis: 0.5 1 3.0 5.0

Odds ratio (OR) for case participants with any pathogenic or likely pathogenic variant (all), those with truncating pathogenic or likely pathogenic variants (PTV), and those with pathogenic missense variants (PMV) for specific cancers in the organ system groupings of cancer *International Classification of Diseases* codes with a significant excess of risk in MyCode. NA indicates not applicable; SAIGE, Scalable and accurate implementation of generalized mixed model.

(adjusted HR, 1.24 [95% CI, 1.13-1.35]; *P* < .001) and UKBB (adjusted HR, 1.17 [95% CI, 1.06-1.30]; *P* = .002). Likewise, case participants in the PTV group were also at higher risk for all cancers tested compared with controls in both MyCode (adjusted HR, 1.30 [95% CI, 1.13-1.50]; *P* < .001) and UKBB (adjusted HR, 1.34 [95% CI, 1.23-1.45]; *P* < .001). There was no significant difference in the penetrance of *CHEK2* PTV vs PMV for cancers in MyCode (univariate HR, 1.06 [95% CI, 0.90-1.25]; *P* = .47) and the UKBB (adjusted HR, 1.15 [95% CI, 1.00-1.33]; *P* = .05).

Figure 3. Organ System–Specific Cancer Risks for Case Participants in UK Biobank

ICD-1O-CM code and cancer type	Participants, No. (%)	OR (95% CI)	Adjusted SAIGE P value
C61 Malignant neoplasm of prostate			
Control	9978 (3.27)	1 [Reference]	NA
Case			
All	170 (5.27)	1.78 (1.48-2.16)	6.68 × 10⁻¹⁰
PMV	55 (4.26)	NA	NA
PTV	107 (5.79)	1.87 (1.47-2.38)	1.27 × 10⁻⁷
C64 Malignant neoplasm of kidney, except renal pelvis			
Control	1635 (0.54)	1 [Reference]	NA
Case			
All	31 (0.96)	1.84 (1.22-2.77)	2.49 × 10⁻³
PMV	11 (0.85)	NA	NA
PTV	20 (1.08)	2.05 (1.23-3.40)	4.28 × 10⁻³
C67 Malignant neoplasm of bladder			
Control	2707 (0.89)	1 [Reference]	NA
Case			
All	45 (1.40)	1.64 (1.17-2.31)	2.66 × 10⁻³
PMV	15 (1.16)	NA	NA
PTV	30 (1.62)	1.88 (1.24-2.85)	2.32 × 10⁻³
C77 Secondary and unspecified malignant neoplasm of lymph nodes			
Control	7493 (2.45)	1 [Reference]	NA
Case			
All	109 (3.38)	1.41 (1.11-1.78)	9.77 × 10⁻⁴
PMV	40 (3.10)	NA	NA
PTV	66 (3.57)	1.50 (1.11-2.03)	4.23 × 10⁻³
C79 Secondary malignant neoplasm of other sites			
Control	6072 (2.00)	1 [Reference]	NA
Case			
All	84 (2.61)	1.34 (1.03-1.76)	.01
PMV	27 (2.09)	NA	NA
PTV	56 (3.03)	1.56 (1.13-2.17)	4.40 × 10⁻³
C81 Hodgkin disease			
Control	414 (0.14)	1 [Reference]	NA
Case			
All	8 (0.25)	1.83 (0.68-4.94)	.02
PMV	0 (0)	NA	NA
PTV	8 (0.43)	3.19 (1.22-8.36)	8.40 × 10⁻³
C83 Diffuse non-Hodgkin lymphoma			
Control	1363 (0.45)	1 [Reference]	NA
Case			
All	26 (0.81)	1.84 (1.06-3.20)	.03
PMV	7 (0.54)	1.27 (0.46-3.55)	>.99
PTV	18 (0.97)	2.19 (1.15-4.17)	.01
C84 Peripheral and cutaneous T-cell lymphomas			
Control	216 (0.07)	1 [Reference]	NA
Case			
All	6 (0.19)	2.68 (0.85-8.45)	.24
PMV	5 (0.39)	5.76 (1.70-19.50)	.02
PTV	1 (0.05)	NA	NA
C85 Other and unspecified types of non-Hodgkin lymphoma			
Control	1533 (0.50)	1 [Reference]	NA
Case			
All	29 (0.90)	1.83 (1.08-3.08)	.02
PMV	8 (0.62)	1.28 (0.49-3.34)	.94
PTV	20 (1.08)	2.18 (1.18-4.00)	.02
C91 Lymphoid leukemia			
Control	921 (0.30)	1 [Reference]	NA
Case			
All	21 (0.65)	2.21 (1.19-4.08)	.01
PMV	9 (0.70)	2.44 (0.99-6.05)	.14
PTV	12 (0.65)	2.16 (0.99-4.74)	.08
C92 Myeloid leukemia			
Control	650 (0.21)	1 [Reference]	NA
Case			
All	20 (0.62)	2.97 (1.58-5.58)	2.17 × 10⁻⁴
PMV	7 (0.54)	2.66 (0.95-7.41)	.13
PTV	13 (0.70)	3.32 (1.56-7.08)	1.22 × 10⁻³
C96 Other and unspecified malignant neoplasms of lymphoid, hematopoietic, and related tissue			
Control	50 (0.02)	1 [Reference]	NA
Case			
All	5 (0.16)	9.47 (2.58-34.80)	1.91 × 10⁻³
PMV	3 (0.23)	NA	NA
PTV	2 (0.11)	NA	NA

Odds ratio (OR) for case participants with any pathogenic or likely pathogenic variant (all), those with truncating pathogenic or likely pathogenic variants (PTV), and those with pathogenic missense variants (PMV) for specific cancers in the organ system groupings of cancer *International Classification of Diseases* codes with a significant excess of risk in Biobank. NA indicates not applicable; SAIGE, Scalable and accurate implementation of generalized mixed model.

All-Cause Mortality

All-cause mortality was significantly increased in the all-variant group of case participants in UKBB, with its consequences observed later in life (age >75 years) (adjusted HR, 1.21 [95% CI, 1.08-1.37]; *P* = .002) but not in MyCode (adjusted HR, 1.09 [95% CI, 0.96-1.24]; *P* = .20) (Figure 4B and Figure 5B). There was no significant difference in all-cause mortality in the PTV and PMV groups in

---

Figure 4. Penetrance of Pathogenic *CHEK2* Variants for Cancer and All-Cause Mortality in MyCode

A Time to cancer
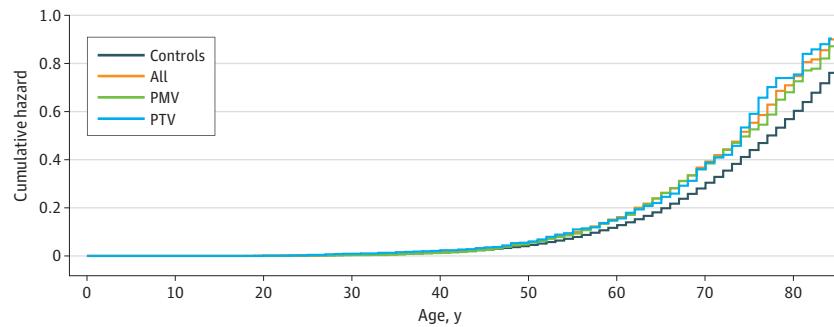
No. at risk								
Controls	148 049	148 047	147 563	138 641	121 024	100 185	71 541	38 520	12 971
All	3021	3021	3015	2837	2497	2073	1446	759	241
PMV	2141	2141	2135	2014	1772	1486	1050	551	180
PTV	881	881	881	824	726	592	403	218	69
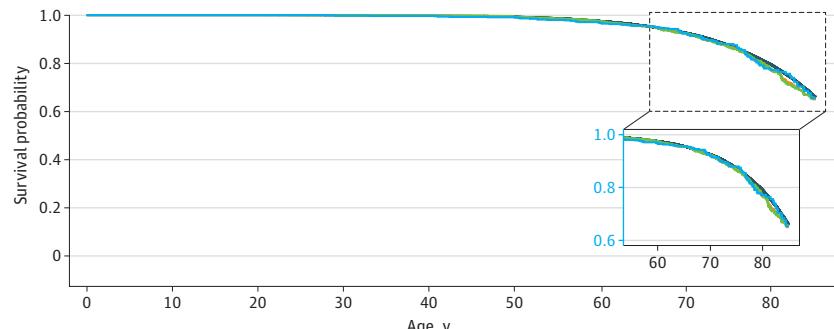
B All-cause mortality

No. at risk								
Controls	151 975	151 973	151 534	143 014	126 032	106 622	80 626	48 585	19 658
All	3133	3133	3128	2958	2631	2257	1702	1041	402
PMV	2221	2221	2216	2099	1866	1616	1233	760	299
PTV	913	913	913	860	766	643	472	288	110
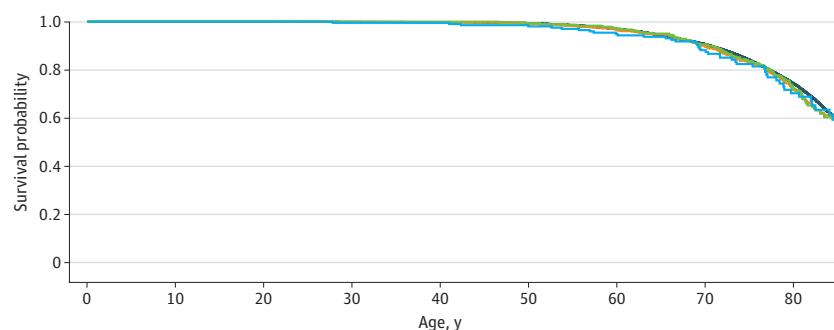
C All-cause mortality for individuals with cancer

No. at risk								
Controls	29 280	29 280	29 279	29 174	28 633	27 276	23 886	17 039	7755
All	732	732	732	726	716	687	599	419	175
PMV	512	512	512	510	505	485	423	298	125
PTV	212	212	212	208	203	195	171	118	49

All includes case participants with any pathogenic or likely pathogenic variant; PMV, those with pathogenic missense variants; and PTV, those with pathogenic truncating variant.

MyCode (adjusted HR, 1.06 [95% CI, 0.80-1.41]; *P* = .67) and UKBB (adjusted HR, 1.24 [95% CI, 0.97-1.60]; *P* = .10).

All-Cause Mortality Among Individuals With Cancer

There was no statistical difference in all-cause mortality among individuals with cancer between case participants in the all-variant group and control participants in both MyCode (adjusted HR, 1.08 [95% CI, 0.90-1.30]; *P* = .43) and the UKBB (adjusted HR, 1.12 [95% CI, 0.98-1.29]; *P* = .11). There were no significant differences between case participants in the PTV and PMV groups in either the MyCode (adjusted HR, 1.20 [95% CI, 0.81-1.78]; *P* = .35) or UKBB (adjusted HR, 1.33 [95% CI, 0.98-1.80]; *P* = .07) cohorts (Figure 4C and Figure 5C).

Discussion

Genomic ascertainment quantifies risk based on genotype (not phenotype) and thus may reduce risk inflation arising from cancer ascertainment (case and family recruitment) by personal and/or family medical history. In this investigation, relatedness-adjusted, Bonferroni-corrected genomic ascertainment of 2 population-based, exome-sequenced, electronic health record–linked cohorts was used to quantify cancer risk and survival from P/LP germline variants in *CHEK2*. Both cohorts had high power to detect elevated risk (OR >2) in all but the rarest cancers. The 3-fold difference in *CHEK2* P variant prevalence in the 2 cohorts (driven especially in missense variation) is explained by the known differences in *CHEK2* allele frequencies in US vs British populations.

Clinically, this investigation confirms the significantly increased risk for breast and prostate cancers (as well as all cancers, collectively), although the observed risk tended to be even lower (OR <2) than previous estimates, especially for those in the PTV group (typically OR >2).[2,19] Interestingly, in neither cohort was a significant excess risk for malignant neoplasms of digestive organs (the majority of which were colorectal cancers) observed for the all-variant, PTV, or PMV groups of case participants (eTable 4 in Supplement 1). Published risk estimates for colorectal cancer from *CHEK2* PTV are more modest (OR of approximately 2) and more conflicting than those for female breast cancer and prostate cancer; higher estimates of risk are driven by studies of multiplex families.[20,21] Published risk estimates for colorectal cancer from *CHEK2* PMV tend to be even lower (OR <2) or not statistically significant.[2,22] Given this, a recent ACMG review and clinical practice guideline on management[2] and current National Comprehensive Cancer Network guidelines (version 3.2024) recommend that *CHEK2* heterozygosity is not clinically actionable for colorectal cancer risk in isolation and to offer surveillance as per family history. In summary, although additional confirmation is needed for breast, prostate, and colorectal cancers, genomic ascertainment showed generally lower (or not significant) risk than previously reported for individuals in the all-variant, PTV, and PMV groups in *CHEK2* variants.

This work provides substantial evidence from both cohorts of significantly increased risk for kidney cancer, bladder cancer, and chronic lymphocytic leukemia (CLL). In this investigation, Bonferroni correction was applied to organ-system groupings and not specific cancer types. Thus, other cancers may be enriched in individuals with heterozygous *CHEK2* variants; eTable 4 in Supplement 1 lists counts of cancer types in control participants and case participants in the all, PTV, and PMV groups. Several publications have reported increased risk of kidney cancer,[22-26] whereas other investigations had nonsignificant findings.[27] As with breast and prostate cancers in this study, the genomic ascertainment used in this study resulted in lower risk estimates (OR <2) for kidney cancer than previous studies and was remarkably consistent across the 2 cohorts. A 2023 ACMG review and clinical guidance for individuals with heterozygous *CHEK2* variants[2] noted a single publication of nonsignificant *CHEK2*-associated bladder cancer[28] but deemed this evidence insufficient to make recommendations; more recent publications have found additional evidence of a *CHEK2*-bladder cancer association.[29,30] Genomic ascertainment in this study revealed similarly increased bladder cancer risk in both cohorts (especially in the PTV groups). Despite the increased

relative risk, the absolute risk must also be considered for these rarer cancers. This is especially important in considering actionability of these findings, given the lack of difference in survival between control and case participants in MyCode and differences late in life in UKBB.

In both cohorts there was significantly elevated risk for lymphoid and hematopoietic neoplasms collectively (C81-C96); across all the subtypes of these malignant neoplasms, only CLL had

Figure 5. Penetrance of Pathogenic *CHEK2* Variants for Cancer and All-Cause Mortality in UK Biobank

A | Time to cancer
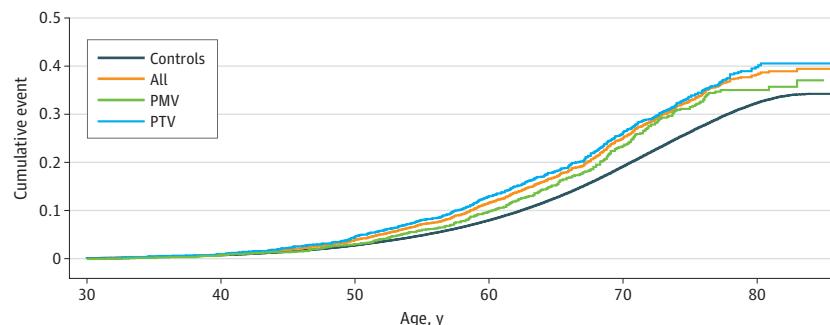
No. at risk

Controls	284 053	302 927	296 571	251 558	151 143	37 197
All	3016	3192	3096	2554	1480	357
PMV	1275	1278	1250	1040	589	132
PTV	1737	1839	1773	1453	855	218

B | All-cause mortality
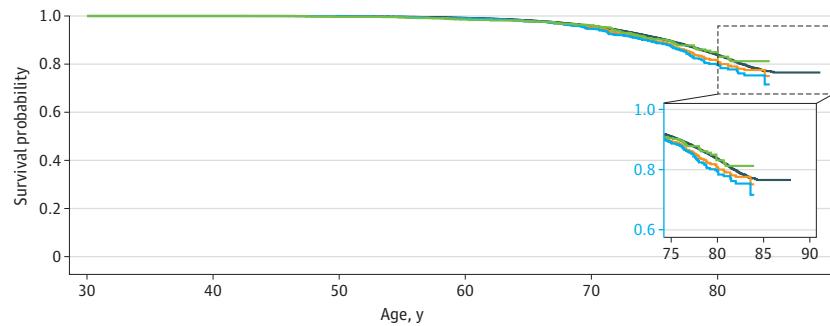
No. at risk

Controls	305 311	305 311	304 854	263 934	170 315	36 963
All	3224	3224	3216	2767	1779	356
PMV	1290	1290	1285	1099	706	131
PTV	1859	1859	1856	1602	1032	215

C | All-cause mortality for individuals with cancer
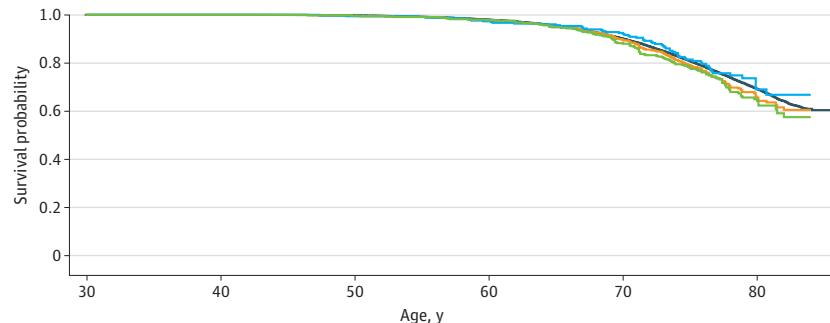
No. at risk

Controls	69 836	69 836	69 669	65 532	49 440	12 539
All	934	934	930	869	638	140
PMV	562	562	560	521	377	91
PTV	344	344	342	321	242	45

All includes case participants with any pathogenic or likely pathogenic variant; PMV, those with pathogenic missense variants; and PTV, those with pathogenic truncating variants.

significantly elevated risk (OR >2) in both cohorts. Reports of increased risk of lymphoid and hematologic malignant neoplasms (especially CLL) in individuals with heterozygous *CHEK2* variants date from 2006[21,31,32] but were conflicting and/or based on highly ascertained families. A 2022 investigation using a phenome-wide association study approach in an earlier version of UKBB reported an excess risk (OR >3) for leukemia and plasma cell neoplasms in individuals with heterozygous *CHEK2* P/LP variants[33]; a 2024 study also using UKBB data found excess risk for Hodgkin lymphoma, diffuse non-Hodgkin lymphoma, and myeloid leukemia in individuals with *CHEK2* PTV.[26] The role of *CHEK2* P variation may provide clues to the etiology of this leukemia; the clinical actionability of these findings should be considered in the context of minimal differences in survival between control and case participants.

A significant excess of malignant neoplasm of thyroid and other endocrine tumors (C73-C75) was observed in MyCode but not UKBB; this was almost entirely driven by thyroid tumors (C73) and, unlike most other associations, by *CHEK2* PMV. Previous studies have been conflicting or limited by small numbers or single-country ascertainment.[22,27,34] Genomic ascertainment of *DICER1*-associated thyroid disease (eg, goiter) also found significant differences in individuals with heterozygous *DICER1* variants (vs control participants) in MyCode but not UKBB and may reflect the different medical cultures in the United States and United Kingdom in approaches to medical imaging of the thyroid.[35] Conversely, there was a significant excess risk of malignant neoplasms of ill-defined, secondary, and unspecified sites (C76-C79) in UKBB but not MyCode.

Numerous other associations have been observed for specific cancers for individuals with heterozygous *CHEK2* variants, including sarcoma, stomach cancer, male breast cancer, melanoma, pancreatic cancer, esophageal cancer, endometrial cancer, and testicular cancer.[2,26] For more common cancers (eg, endometrial, skin), there was no evidence of association for these in either cohort. For some rarer cancers (male breast, testicular), the 2 cohorts were likely underpowered (eFigure 5 in Supplement 1); for others (sarcoma, stomach), there may be both a power issue and a survival bias in ascertainment given the aggressive nature of these cancers.

Limitations

There are limitations to these retrospective analyses. Participants in MyCode and UKBB are predominantly of European ancestry. Copy-number variants in *CHEK2* were not evaluated due to limited data availability in UKBB. Enrollment in the 2 cohorts was subject to ascertainment biases, as individuals with conditions leading to death or disabilities would be less likely to participate. The healthy volunteer bias (compared with the UK population) of the UKBB has been documented.[36] Absolute risk was not quantified.

Conclusions

In this case-control study, we evaluated cancer risk and survival in individuals with heterozygous *CHEK2* variants using the novel genome-first approach in 2 well-powered cohorts. Pathogenic germline *CHEK2* subgroups of all variants, PTV, and PMV were common in European populations. In addition to breast and prostate risk, we found evidence in both cohorts of associations with kidney and bladder cancers and CLL that may provide clues to etiology. With genomic ascertainment, the conferred excess cancer risk was low (OR <2). This has clinical and counseling implications for individuals ascertained this way (vs with a family history of cancer). In addition, the lack of significant difference between case and control participants in all-cause mortality in individuals with cancer suggests that germline *CHEK2*-associated cancer was not clinically more aggressive than non–*CHEK2*-associated cancer. The degree of risk from PTV and PMV overlap considerably, with risk of PMV generally lower. However, cancer penetrance, all-cause mortality, and all-cause mortality in individuals with cancer was not significantly different between PMV and PTV, suggesting that clinical differences between these variant types are less relevant.

Author Affiliations: Clinical Genetics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, Rockville, Maryland (S. Y. Kim, J. Kim, Ramos, Stewart); Louisiana State University Health Sciences Center, New Orleans, Louisiana (S. Y. Kim); now with Pennsylvania State University, College of Health and Human Development, State College (Ramos); Department of Genomic Health, Geisinger, Danville, Pennsylvania (Haley, Smelser, Rao, Mirshahi, Carey); Division of Translational Medicine and Human Genetics, Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA and (Nathanson); Abramson Cancer Center, Perelman School of Medicine, University of Pennsylvania, Philadelphia (Nathanson); Biostatistics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Rockville, Maryland (Graubard, Katki).

Geisinger Regeneron DiscovEHR collaboration contributors who have been critical in the generation of the data used for this study.

REFERENCES

1. Stolz A, Ertych N, Bastians H. Tumor suppressor CHK2: regulator of DNA damage response and mediator of chromosomal stability. *Clin Cancer Res*. 2011;17(3):401-405. doi:10.1158/1078-0432.CCR-10-1215

2. Hanson H, Astiazaran-Symonds E, Amendola LM, et al; ACMG Professional Practices and Guidelines Committee. Electronic address: documents@acmg.net. Management of individuals with germline pathogenic/likely pathogenic variants in CHEK2: a clinical practice resource of the American College of Medical Genetics and Genomics (ACMG). *Genet Med*. 2023;25(10):100870. doi:10.1016/j.gim.2023.100870

3. Cybulski C, Górski B, Huzarski T, et al. CHEK2 is a multiorgan cancer susceptibility gene. *Am J Hum Genet*. 2004;75(6):1131-1135. doi:10.1086/426403

4. Dorling L, Carvalho S, Allen J, et al; NBCS Collaborators; kConFab Investigators; SGBCC Investigators. Breast cancer risks associated with missense variants in breast cancer susceptibility genes. *Genome Med*. 2022;14(1):51. doi:10.1186/s13073-022-01052-8

5. Wilczewski CM, Obasohan J, Paschall JE, et al. Genotype first: clinical genomics research through a reverse phenotyping approach. *Am J Hum Genet*. 2023;110(1):3-12. doi:10.1016/j.ajhg.2022.12.004

6. Carey DJ, Fetterolf SN, Davis FD, et al. The Geisinger MyCode community health initiative: an electronic health record-linked biobank for precision medicine research. *Genet Med*. 2016;18(9):906-913. doi:10.1038/gim.2015.187

7. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature*. 2018;562(7726):203-209. doi:10.1038/s41586-018-0579-z

8. Conroy MC, Lacey B, Bešević J, et al. UK Biobank: a globally important resource for cancer research. *Br J Cancer*. 2023;128(4):519-527. doi:10.1038/s41416-022-02053-5

9. Dewey FE, Murray MF, Overton JD, et al. Distribution and clinical impact of functional variants in 50,726 whole-exome sequences from the DiscovEHR study. *Science*. 2016;354(6319):aaf6814. doi:10.1126/science.aaf6814

10. Cingolani P, Platts A, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin)*. 2012;6(2):80-92. doi:10.4161/fly.19695

11. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010;38(16):e164. doi:10.1093/nar/gkq603

12. Landrum MJ, Lee JM, Benson M, et al. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res*. 2018;46(D1):D1062-D1067. doi:10.1093/nar/gkx1153

13. Li Q, Wang K. InterVar: clinical interpretation of genetic variants by the 2015 ACMG-AMP guidelines. *Am J Hum Genet*. 2017;100(2):267-280. doi:10.1016/j.ajhg.2017.01.004

14. Richards S, Aziz N, Bale S, et al; ACMG Laboratory Quality Assurance Committee. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*. 2015;17(5):405-424. doi:10.1038/gim.2015.30

15. Chow SC, Shao J, Wang H. *Sample Size Calculations in Clinical Research*. 2nd ed. Chapman & Hall/CRC; 2008.

16. All of Us Research Hub. Genomic variants. Accessed November 10, 2025. https://databrowser.researchallofus.org/snvsindels

17. Sun KY, Bai X, Chen S, et al; Regeneron Genetics Center; RGC-ME Cohort Partners. A deep catalogue of protein-coding variation in 983,578 individuals. *Nature*. 2024;631(8021):583-592. doi:10.1038/s41586-024-07556-0

18. Regeneron Genetics Center. Million Exome Variant Browser. Accessed November 10, 2025. https://rgc-research.regeneron.com/me/license-and-terms-of-use

19. Rowlands CF, Allen S, Balmaña J, et al. Population-based germline breast cancer gene association studies and meta-analysis to inform wider mainstream testing. *Ann Oncol*. 2024;35(10):892-901. doi:10.1016/j.annonc.2024.07.244

20. Meijers-Heijboer H, Wijnen J, Vasen H, et al. The CHEK2 1100delC mutation identifies families with a hereditary breast and colorectal cancer phenotype. *Am J Hum Genet*. 2003;72(5):1308-1314. doi:10.1086/375121

21. Schreurs MAC, Schmidt MK, Hollestelle A, et al; Hereditary Breast and Ovarian Cancer Research Group Netherlands. Cancer risks for other sites in addition to breast in CHEK2 c.1100delC families. *Genet Med*. 2024;26(9):101171. doi:10.1016/j.gim.2024.101171

22. Bychkovsky BL, Agaoglu NB, Horton C, et al. Differences in cancer phenotypes among frequent CHEK2 variants and implications for clinical care—checking CHEK2. *JAMA Oncol*. 2022;8(11):1598-1606. doi:10.1001/jamaoncol.2022.4071

23. Weischer M, Bojesen SE, Tybjaerg-Hansen A, Axelsson CK, Nordestgaard BG. Increased risk of breast cancer associated with CHEK2*1100delC. *J Clin Oncol*. 2007;25(1):57-63. doi:10.1200/JCO.2005.05.5160

24. Carlo MI, Mukherjee S, Mandelker D, et al. Prevalence of germline mutations in cancer susceptibility genes in patients with advanced renal cell carcinoma. *JAMA Oncol*. 2018;4(9):1228-1235. doi:10.1001/jamaoncol.2018.1986

25. Han SH, Camp SY, Chu H, et al. Integrative analysis of germline rare variants in clear and non-clear cell renal cell carcinoma. *Eur Urol Open Sci*. 2024;62:107-122. doi:10.1016/j.euros.2024.02.006

26. Mukhtar TK, Wilcox N, Dennis J, et al. Protein-truncating and rare missense variants in *ATM* and *CHEK2* and associations with cancer in UK Biobank whole-exome sequence data. *J Med Genet*. 2024;61(11):1016-1022. doi:10.1136/jmg-2024-110127

27. Näslund-Koch C, Nordestgaard BG, Bojesen SE. Increased risk for other cancers in addition to breast cancer for CHEK2*1100delC heterozygotes estimated from the Copenhagen General Population Study. *J Clin Oncol*. 2016;34(11):1208-1216. doi:10.1200/JCO.2015.63.3594

28. Pemov A, Wegman-Ostrosky T, Kim J, et al. Identification of genetic risk factors for familial urinary bladder cancer: an exome sequencing study. *JCO Precis Oncol*. Published online December 22, 2021. doi:10.1200/PO.21.00115

29. Mian A, Wei J, Shi Z, et al. Systematic review of reported association studies of monogenic genes and bladder cancer risk and confirmation analysis in a large population cohort. *BJUI Compass*. 2022;4(2):156-163. doi:10.1002/bco2.206

30. Yang Y, Zhang G, Hu C, et al. The germline mutational landscape of genitourinary cancers and its indication for prognosis and risk. *BMC Urol*. 2022;22(1):196. doi:10.1186/s12894-022-01141-1

31. Rudd MF, Sellick GS, Webb EL, Catovsky D, Houlston RS. Variants in the ATM-BRCA2-CHEK2 axis predispose to chronic lymphocytic leukemia. *Blood*. 2006;108(2):638-644. doi:10.1182/blood-2005-12-5022

32. Stubbins RJ, Korotev S, Godley LA. Germline CHEK2 and ATM variants in myeloid and other hematopoietic malignancies. *Curr Hematol Malig Rep*. 2022;17(4):94-104. doi:10.1007/s11899-022-00663-7

33. Zeng C, Bastarache LA, Tao R, et al. Association of pathogenic variants in hereditary cancer genes with multiple diseases. *JAMA Oncol*. 2022;8(6):835-844. doi:10.1001/jamaoncol.2022.0373

34. Kaczmarek-Ryś M, Ziemnicka K, Hryhorowicz ST, et al. The c.470 T > C CHEK2 missense variant increases the risk of differentiated thyroid carcinoma in the Great Poland population. *Hered Cancer Clin Pract*. 2015;13(1):8. doi:10.1186/s13053-015-0030-5

35. Kim J, Haley J, Hatton JN, et al. A genome-first approach to characterize DICER1 pathogenic variant prevalence, penetrance and cancer, thyroid, and other phenotypes in 2 population-scale cohorts. *Genet Med Open*. Published online April 11, 2024. doi:10.1016/j.gimo.2024.101846

36. Fry A, Littlejohns TJ, Sudlow C, et al. Comparison of sociodemographic and health-related characteristics of UK Biobank participants with those of the general population. *Am J Epidemiol*. 2017;186(9):1026-1034. doi:10.1093/aje/kwx246

SUPPLEMENT 1.
eMethods.
eFigure 1. Odds Ratio for All, PTV, and PMV Groups of Case Participants for Organ System Groupings of Cancer *ICD* Codes in MyCode
eFigure 2. Odds Ratio for All, PTV, and PMV Groups of Case Participants for Organ System Groupings of Cancer ICD Codes in UK Biobank
eFigure 3. Odds Ratio for All, PTV, and PMV Groups of Case Participants for All Specific Cancers in the Organ System Groupings of Cancer ICD Codes in MyCode
eFigure 4. Odds Ratio for All, PTV, and PMV Groups of Case Participants for All Specific Cancers in the Organ System Groupings of Cancer ICD Codes in UK Biobank
eFigure 5. Power as a Function of Risk (Odds Ratio) in MyCode and UK Biobank for a Range of Cancer Rates

SUPPLEMENT 2.
eTable 1. List of all variants found in the study
eTable 2: Frequency of six common *CHEK2* variants in different biobanks
eTable 3. Demographics of *CHEK2* heterozygotes vs controls

eTable 4. Case counts and percentages for PMV, PTV and All cohorts and fold-enrichment (vs controls) for each of the ICD10 diagnostic codes in MyCode and UK Biobank

eTable 5. Prevalence of All Variants, PTVs, and PMVs in *CHEK2* Among Adults in UK Biobank and Geisinger MyCode

SUPPLEMENT 3.

Nonauthor Collaborators

SUPPLEMENT 4.

Data Sharing Statement