Research Article

Cancer Epidemiology, Biomarkers & Prevention

Analysis of Heritability and Genetic Architecture of Pancreatic Cancer: A PanC4 Study

Fei Chen[1], Erica J. Childs[2], Evelina Mocci[2], Paige Bracci[3], Steven Gallinger[4], Donghui Li[5], Rachel E. Neale[6], Sara H. Olson[7], Ghislaine Scelo[8], William R. Bamlet[9], Amanda L. Blackford[2], Michael Borges[10], Paul Brennan[8], Kari G. Chaffee[9], Priya Duggal[1], Manal J. Hassan[5], Elizabeth A. Holly[3], Rayjean J. Hung[11], Michael G. Goggins[10], Robert C. Kurtz[12], Ann L. Oberg[9], Irene Orlow[7], Herbert Yu[13], Gloria M. Petersen[9], Harvey A. Risch[14], and Alison P. Klein[1,2,10]

Abstract

Background: Pancreatic cancer is the fourth-leading cause of cancer death in both men and women in the United States. The currently identified common susceptibility loci account for a small fraction of estimated heritability. We sought to estimate overall heritability of pancreatic cancer and partition the heritability by variant frequencies and functional annotations.

Methods: Analysis using the genome-based restricted maximum likelihood method (GREML) was conducted on Pancreatic Cancer Case-Control Consortium (PanC4) genome-wide association study (GWAS) data from 3,568 pancreatic cancer cases and 3,363 controls of European Ancestry.

Results: Applying linkage disequilibrium- and minor allele frequency-stratified GREML (GREML-LDMS) method to imputed GWAS data, we estimated the overall heritability of pancreatic cancer to be 21.2% (SE = 4.8%). Across the functional groups (intronic, intergenic, coding, and regulatory variants), intronic variants account for most of the estimated heritability (12.4%). Previously identified GWAS loci explained 4.1% of the total phenotypic variation of pancreatic cancer. Mutations in hereditary pancreatic cancer susceptibility genes are present in 4% to 10% of patients with pancreatic cancer, yet our GREML-LDMS results suggested these regions explain only 0.4% of total phenotypic variance for pancreatic cancer.

Conclusions: Although higher than previous studies, our estimated 21.2% overall heritability may still be downwardly biased due to the inherent limitation that the contribution of rare variants in genes with a substantive overall impact on disease are not captured when applying these commonly used methods to imputed GWAS data.

Impact: Our work demonstrated the importance of rare and common variants in pancreatic cancer risk.

Introduction

Pancreatic cancer is one of the most lethal malignant neoplasms across the world. The highest incidence and mortality rates are found in North America and Western Europe, followed by other more developed regions (1). Pancreatic cancer is currently the fourth-leading cause of cancer death in both men and women in the United States, responsible for an estimated 44,330 deaths in 2018 (2). By 2030, pancreatic cancer is predicted to become the second most common cause of cancer mortality (3). Up to 10% of patients with pancreatic cancer report having a first-degree relative (FDR) affected by the disease, and up to 10% of all newly diagnosed patients with pancreatic cancer harbor a germline mutation in a hereditary pancreatic cancer susceptibility gene (4–6).

Although only a handful of studies have examined the heritability of pancreatic cancer, a large population-based twin study in European countries estimated the heritability of pancreatic cancer to be 36% (95% CI, 0%–53%; ref. 7). Inherited genetic mutations

[1]Department of Epidemiology, Johns Hopkins School of Public Health, Baltimore, Maryland. [2]Department of Oncology, Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins School of Medicine, Baltimore, Maryland. [3]Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, California. [4]Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Ontario, Canada. [5]Department of Gastrointestinal Medical Oncology, University of Texas MD Anderson Cancer Center, Houston, Texas. [6]Department of Population Health, QIMR Berghofer Medical Research Institute, Queensland, Australia. [7]Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, New York. [8]International Agency for Research on Cancer (IARC), Lyon, France. [9]Department of Health Sciences Research, Mayo Clinic College of Medicine, Rochester, Minnesota. [10]Department of Pathology, Sol Goldman Pancreatic Cancer Research Center, Johns Hopkins School of Medicine, Baltimore, Maryland. [11]Lunenfeld-Tanenbaum Research Institute of Mount Sinai Hospital, Toronto, Ontario, Canada. [12]Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, New York. [13]Epidemiology Program, University of Hawaii Cancer Center, Honolulu, Hawaii. [14]Department of Chronic Disease Epidemiology, Yale School of Public Health, New Haven, Connecticut.

Note: Supplementary data for this article are available at Cancer Epidemiology, Biomarkers & Prevention Online (http://cebp.aacrjournals.org/).

Corresponding Author: Alison P. Klein, Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins School of Medicine, 354 CRB2, 1550 Orleans Street, Baltimore, MD 21231. Phone: 410-955-3511; Fax: 410-614-0671; E-mail: aklein1@jhmi.edu

Cancer Epidemiol Biomarkers Prev 2019;28:1238–45

doi: 10.1158/1055-9965.EPI-18-1235

©2019 American Association for Cancer Research.

Downloaded from http://aacrjournals.org/cebp/article-pdf/28/7/1238/2285991/1238.pdf by guest on 03 November 2024

in 11 genes including *BRCA2*, *ATM*, *CDKN2A*, *PALB2*, *BRCA1*, *PRSS1*, *STK11*, *MLH1*, *MSH2*, *MHS6*, and *PMS2* have been associated with an increased risk of pancreatic cancer. Overall, 8% to 30% of patients with familial pancreatic cancer (FPC; refs. 8–11) and 3% to 10% of unselected pancreatic cancer cases (4–6) harbor a deleterious mutation in one of these 11 genes, demonstrating the important role of these genes in the pancreatic cancer susceptibility. Recent genome-wide association studies (GWAS) in European (12–17) and Asian (18, 19) populations have identified 26 independent common susceptibility loci for pancreatic cancer. Despite the large sample sizes of these GWAS, the identified common susceptibility loci together explain <5% of the total phenotypic variation (pancreatic cancer/not pancreatic cancer) for pancreatic cancer (20, 12, 21). Comparing this with the family-based estimate of heritability (36%; ref. 7), it appears that a large proportion of heritability is unexplained, highlighting the so-called "missing heritability" problem. Except for some conditions, such as age-related macular degeneration in which heritability is substantially explained by a small number of common variants of large effect, for most complex traits or diseases the proportion of heritability explained remains small despite a large number of identified variants (22). Potential sources of missing heritability are thought to be either rare variants not well tagged by GWAS arrays or the common variants that have not yet reached statistical significance in prior GWAS studies. Given that genetic architecture varies across traits, the sources of missing heritability are likely variable as well.

To better understand the sources of missing heritability, approaches including the genomic relatedness-based restricted maximum-likelihood (GREML) were developed to quantify the cumulative effects of causal variants in populations of unrelated individuals (23). Heritability estimation using pedigree data is a foundation of genetic epidemiologic studies. However, given the late age of onset and rarity of pancreatic cancer, there is limited power even in studies using the largest registries to estimate the heritability of pancreatic cancer (24). In addition, it has been suggested that pedigree-based heritability estimates can be upwardly biased due to the sharing of nongenetic factors among pedigree members (25, 26). In contrast, newer methods such as GREML, that estimates genetic relationships using genome-wide array, are thought to overcome this bias. The early version of GREML, single-component GREML (GREML-SC), has been widely applied in GWAS to estimate heritability. In pancreatic cancer GWAS, heritability using this approach was reported to range from 9.8% to 18% (12, 21, 20).

However, despite its wide application in GWAS studies, heritability estimated from GREML-SC is known to be biased (27). To overcome this bias, a multicomponent GREML approach was developed which allows for stratification on minor allele frequency (MAF) and linkage disequilibrium (LD). The LD- and MAF-stratified GREML (GREML-LDMS) has been shown to produce more valid estimates of heritability across different simulated scenarios (27, 28). The multicomponent GREML approach not only provides less biased heritability estimates but also allows for the estimation of heritability components from different variant sets.

The goal of this study was to understand the genetic architecture of pancreatic cancer by applying a multicomponent approach to GWAS array data after imputation.

Materials and Methods

Study participants

The data used in this study were obtained from the Pancreatic Cancer Case Control Consortium (PanC4) GWAS, which comprises 9 hospital-based or population-based case–control studies (http://panc4.org; ref. 12). Participating sites include Johns Hopkins Hospital (Baltimore, MD), Mayo Clinic (Rochester, MN), MD Anderson Cancer Center (Houston, TX), Memorial Sloan-Kettering Cancer Center (New York, NY), Yale University (New Haven, CT), University of Toronto (Toronto, Canada), University of California San Diego (San Diego, CA), Queensland (Queensland, Australia), and International Agency for Research on Cancer (Central Europe). Cases were defined as individuals with adenocarcinoma of the pancreas and controls were individuals without a diagnosis of pancreatic cancer sampled from the general population or hospital catchment area as described previously (13). In brief, the mean age of cases was 64.7 years compared with 63.1 years in controls, 58% of the participants were male and 95% reported European Ancestry. This study was reviewed and approved by the Institutional Review Board of the Johns Hopkins University School of Medicine, and of each participating institution. Informed consent was obtained from all participants in this study.

Genotyping, imputation, and quality control

A total of 7,956 PanC4 participants were genotyped with the IlluminaHumanOmniExpressExome-8v1 array; additional variants were imputed using IMPUTE v2 (29) to the 1000 Genomes (phase III, v3; ref. 30) reference panel. Details on the genotyping and imputation have been described previously (13). After imputation, the genotype imputation probabilities were converted to hard genotype calls using PLINK (the genotype with the highest probability was the hard genotype unless the difference between the highest two probabilities is less than 0.1, in which case genotypes were set to missing; ref. 31). The following quality control filters were applied to the 81,671,345 autosomal variants in accordance with the GREML recommendations in which we: (i) removed 372 known non-European samples, (ii) dropped variants with INFO score less than 0.50, (iii) dropped variants that failed Hardy–Weinberg equilibrium exact test at $P < 10^{-6}$, and (iv) dropped variants with a minor allele count of less than 5 (equivalent to a MAF < 0.0003). After quality control checks, 1.9% variants with missingness greater than 5% were excluded, and 60% of the variants were dropped due to being monomorphic. As GREML is sensitive to cryptic relatedness, genetic relatedness was determined using 99,138 common (MAF > 0.05) and independent (pairwise $r^2 > 0.20$) variants directly genotyped in the dataset. At a relatedness cutoff of 0.025, 653 distantly related individuals were excluded. The final dataset contained 6,931 samples and 16,184,129 variants (Supplementary Fig. S1). Annotation of the variants was obtained from ANNOVAR (32). The functional predictions were derived from the NCBI Reference Sequence Database (33).

Statistical analysis

Estimation of heritability using GREML-LDMS. The proportion of phenotypic variation explained by all imputed variants was estimated in a GREML-LDMS model. Variants were stratified into 2 MAF bins (MAF < 0.01 and MAF $\geq$ 0.01), as well as 2 LD groups as above or below the median regional LD score. A sliding window

Downloaded from http://aacrjournals.org/cebp/article-pdf/28/7/1238/2285991/1238.pdf by guest on 03 November 2024

method was used to determine the regional LD score for each variant (28). The genetic relationship matrix (GRM) from each MAF-LD stratum were calculated and fitted jointly in a mixed linear model using the average information approach for variance estimation. Estimates of variance were transformed from the observed 0 to 1 scale to the unobserved continuous "liability" scale using a probit transformation (34). A disease prevalence of 0.0149 was specified, which corresponds to the lifetime risk of being diagnosed with cancer of the pancreas for U.S. whites in the 2009 to 2011 SEER report (35). All analyses were adjusted for age, sex, and the first 20 principal components. The variance in the liability scale was reported along with its SEs. Potential bias in the estimated heritability due to residual population stratification and/or relatedness was quantified by comparing the variance explained by individual chromosomes in a separate analysis to that in a joint analysis, as previously described (36). For all analyses, SEs of the summed variance were calculated from the sample variance/covariance matrix using the delta method.

Genomic partitioning by chromosome. To determine the variance captured by each autosomal chromosome, the variants in 4 MAF-LD groups were further allocated to 22 autosomal chromosomes, resulting in 88 MAF–LD–chromosome strata. The Fisher scoring approach was used in this analysis for variance estimation. The variance captured by each chromosome was aggregated from the variance due to 4 MAF–LD groups allocated to each chromosome. Linear regression was performed to assess the correlations between variance explained by an individual chromosome and the length of the chromosome, defined as the total number of variants in the chromosome.

Genomic partitioning by MAF. To improve the resolution in the MAF distribution of causal variants, variants were binned into 6 MAF categories: $0.0003 \leq \text{MAF} < 0.01$, $0.01 \leq \text{MAF} < 0.10$, $0.10 \leq \text{MAF} < 0.20$, $0.20 \leq \text{MAF} < 0.30$, $0.30 \leq \text{MAF} < 0.40$, and $\text{MAF} \geq 0.40$. Variants in each MAF category were then stratified by their regional LD score (above vs. below median LD) as done previously, resulting in 12 MAF-LD strata. GRMs calculated from each stratum were fitted jointly in a mixed linear model. The variance captured by each MAF category was aggregated from the variance due to 2 LD strata within the MAF category.

Genomic partitioning by functional annotations. Imputed variants were categorized in 4 functional groups: coding (including exonic and splicing variants); intergenic; intronic; and regulatory [including noncoding RNA, variants in untranslated regions (UTR), and upstream/downstream variants]. Variants in each of the 4 functional groups were further stratified into 2 MAF categories and 2 LD groups as in previous analysis. In the joint analysis of all functional groups, the variance explained by each functional category was summed from the variance due to 4 MAF–LD strata within the functional category.

Contribution of GWAS loci. A total of 26 loci previously identified by GWAS have reported to be significantly associated with pancreatic cancer risk at the genome-wide level (12–19). The index SNP or the variants with the strongest LD (pairwise r^2 in 1000 Genomes EUR population) to the index SNP were included in the estimate to capture the GWAS signals. Then all variants within ±250 kb of the index SNP were grouped together with the index SNP as a single genetic component. The remaining variants across the genome were stratified into 2 MAF categories and 2 LD groups as in previous analyses. The variance explained by the GWAS loci was estimated by fitting 5 GRMs jointly in a mixed linear model.

Contribution of established FPC genes. To evaluate the contribution of established FPC genes, all variants located within ±50 kb

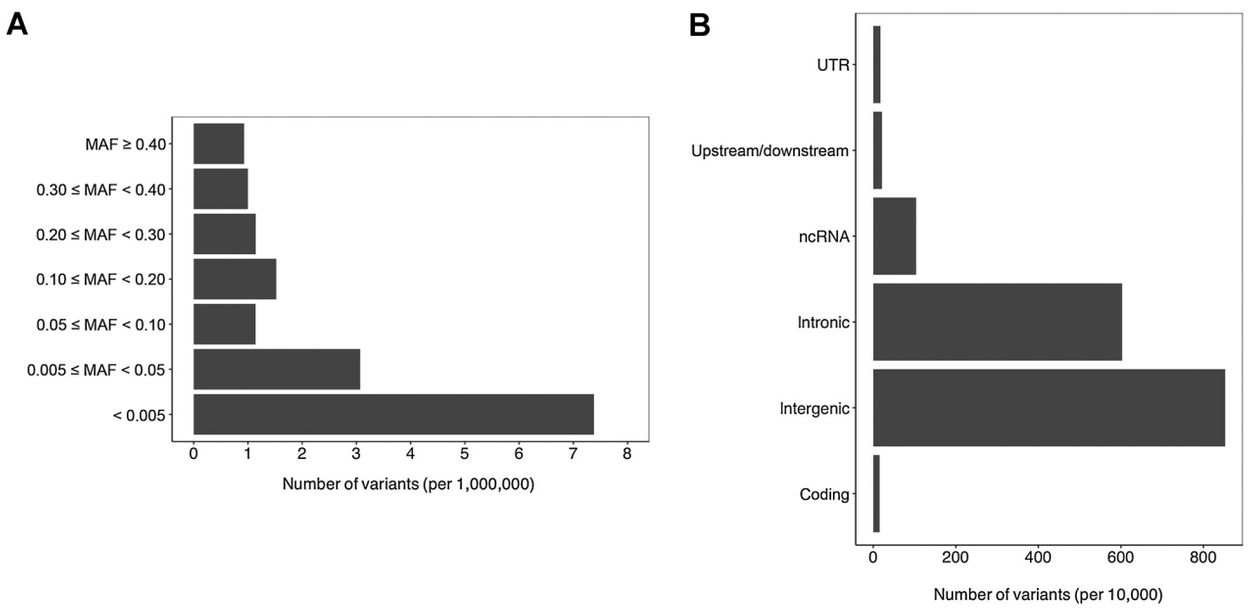

Figure 1.

MAF and functional annotation of PanC4 imputed variants. **A,** MAF distribution of imputed variants passed all quality control filters showed that majority of these variants had an MAF < 0.05. **B,** Imputed variants were annotated into 6 functional groups by ANNOVA, among which intergenic (52.7%) and intronic (37.2%) variants were the 2 largest groups.

Downloaded from http://aacrjournals.org/cebp/article-pdf/28/7/1238/2285991/1238.pdf by guest on 03 November 2024

Table 1. Estimates of variance explained by imputed variants from GREML-LDMS analysis

	Above mean LD		Below mean LD				
	Est	SE	Est	SE	Row sum	SE[a]	No. variants (%)
MAF < 0.01	0.052	0.037	0	0.051	0.052	0.036	8,354,405 (51.6%)
MAF ≥ 0.01	0.014	0.017	0.146	0.032	0.160	0.035	7,829,724 (48.4%)
Column sum	0.067	0.040	0.146	0.058			
Total sum	0.212	0.048					
No. variants (%)	8,092,536 (50%)	8,091,593 (50%)					

Abbreviation: Est, estimate.
[a]SE for row sum and column sum was calculated from variance/covariance matrix.

of gene boundaries (3′ UTR to 5′ UTR) of these genes were used to calculate a single GRM. The remaining variants across the genome were stratified into 2 MAF categories and 2 LD groups as described previously. The variance explained by these 11 genes was estimated by fitting 5 GRMs jointly in a mixed linear model.

Results

The final analytical population included 3,568 pancreatic cancer cases and 3,363 controls, all of whom were of European ancestry and ages 40 years or older. Cases and controls were similar in sex and age distributions. Fig. 1A shows the distribution of MAFs in the final dataset containing 16,184,129 variants. The majority of the variants have a MAF < 5%. The remaining variants are evenly distributed across the MAF frequency categories. More than half of the variants in the final dataset are intergenic (52.7%) or intronic (37.2%). About 1% of the variants were located in coding regions (Fig. 1B).

In PanC4 study, imputed variants explained in total of 21.2% (SE = 4.8%) phenotypic variation for pancreatic cancer (Table 1). We assessed the potential inflation due to residual population stratification and/or cryptic relatedness by examining heritability on each individual chromosome, and obtained an estimate of 0.31%, suggesting minimal inflation.

Genomic partitioning of the estimated heritability can provide valuable insights on the underlying genetic architecture of the disease. The estimated variance associated with each autosomal chromosome is shown in Fig. 2. Chromosome 9 accounted for the largest proportion of genetic variation ($h^2 = 2.3\%$, SE = 1.6%), followed by chromosome 7 ($h^2 = 2.1\%$, SE = 1.8%).

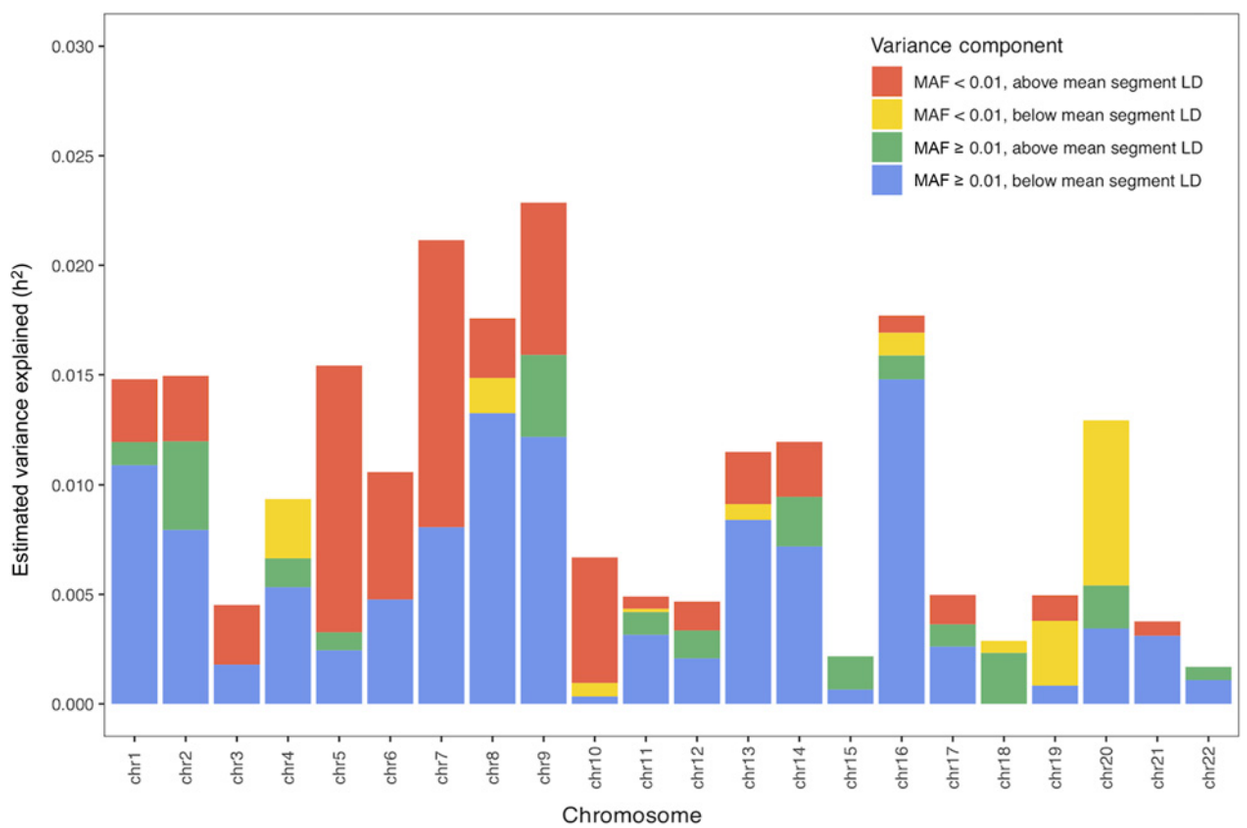

Figure 2.
Estimated variance explained by imputed variants on individual chromosome stratified by MAF and LD. Variants on each chromosome were stratified into 2 MAF categories and 2 LD groups. The estimated variance associated with individual chromosome was aggregated from the variance explained by 4 MAF–LD groups. This analysis ranks chromosome 9, 7, 16, 8, 5, 2, and 1 as top contributors to the estimated heritability.

Downloaded from http://aacrjournals.org/cebp/article-pdf/28/7/1238/2285991/1238.pdf by guest on 03 November 2024

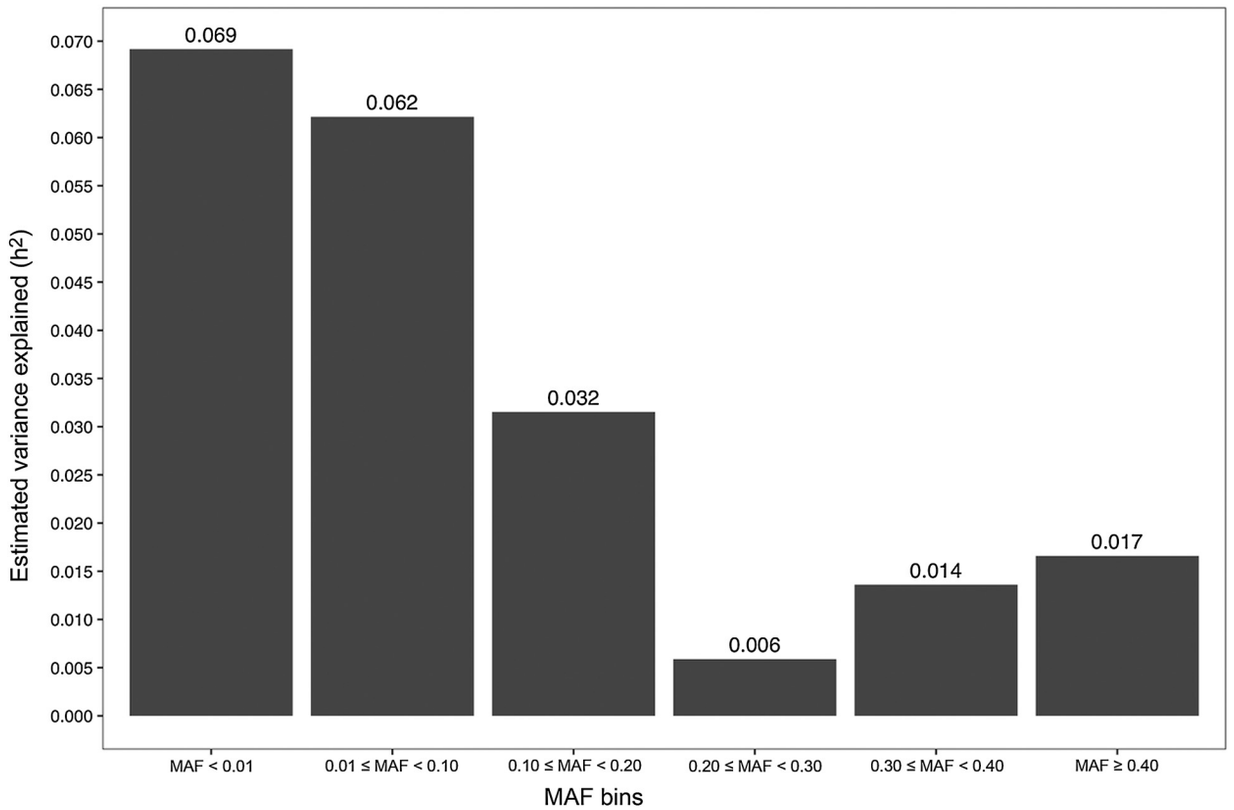

Figure 3.

Estimated variance explained by imputed variants stratified by MAF. Variants were stratified into 6 MAF categories: <0.01, 0.01–0.10, 0.10–0.20, 0.20–0.30, 0.30–0.40, and ≥0.40. Across the MAF categories, rare variants with MAF < 0.01 accounts for the most variance, followed by variants with MAF ranged from 0.01 to 0.10.

Chromosomes 8 ($h^2 = 1.8\%$, SE = 1.7%), 16 ($h^2 = 1.8\%$, SE = 1.4%), 5 ($h^2 = 1.5\%$, SE = 1.9%), 2 ($h^2 = 1.5\%$, SE = 2.0%), and 1 ($h^2 = 1.5\%$, SE = 2.0%). Common susceptibility loci for pancreatic cancer have been identified in GWAS studies on each of these chromosomes. Regression of the variance explained by individual chromosomes on the length of the chromosome found no correlations (Supplementary Fig. S2).

Partitioning of the estimated heritability by 6 MAF categories found a substantial amount of genetic variation for pancreatic cancer attributed to rare variants, with $h^2 = 6.9\%$ (SE = 3.8%) for variants with MAF < 0.01, corresponding to one-third of the estimated heritability (Fig. 3). Variants with 0.01 ≤ MAF < 0.10 explain a comparable amount of variance for pancreatic cancer ($h^2 = 6.2\%$, SE = 3.1%).

In the genomic partitioning by functional groups, intronic and intergenic variants account for 12.4% (SE = 6.6%) and 6.0% (SE = 6.8%) of phenotypic variance for pancreatic cancer, respectively (Table 2). Coding variants, including exonic and splicing variants, despite being the smallest functional group, explained 1.0% (SE = 3.9%) of the phenotypic variance for pancreatic cancer. The remaining 1.8% variance (SE = 4.5%) was attributed to variants in regulatory regions (UTR, ncRNA, and upstream/downstream).

Of the 26 common susceptibility loci reported in GWAS, 23 index SNPs were available in our dataset. For the 3 GWAS loci whose index SNP was not available in our dataset, including rs2736098 on chromosome 5p13.33 (*TERT-CLPTM1L*), rs10094872 on chromosome 8q24.21 (*MYC*), and rs4795218

Table 2. Estimates of variance explained by imputed variants in 4 functional groups

		Above mean LD		Below mean LD						
		Est	SE	Est	SE	Row sum	SE	No. variants (%)	Subcategory sum	SE
Coding	MAF < 0.01	0.002	0.022	0.004	0.029	0.005	0.036	103,133 (0.6%)	0.010	0.039
	MAF ≥ 0.01	0.005	0.009	0	0.014	0.005	0.017	50,753 (0.3%)		
Intergenic	MAF < 0.01	0.033	0.036	0	0.059	0.033	0.063	4,305,707 (26.6%)	0.060	0.068
	MAF ≥ 0.01	0	0.015	0.027	0.026	0.027	0.030	4,220,706 (26.1%)		
Intronic	MAF < 0.01	0.042	0.037	0	0.055	0.042	0.061	3,173,417 (19.6%)	0.124	0.066
	MAF ≥ 0.01	0.009	0.014	0.074	0.025	0.082	0.028	2,856,226 (17.6%)		
Regulatory	MAF < 0.01	0	0.024	0	0.036	0	0.043	746,866 (4.6%)	0.018	0.045
	MAF ≥ 0.01	0.0002	0.009	0.018	0.014	0.018	0.016	676,634 (4.2%)		

Abbreviation: Est, estimate.

Downloaded from http://aacrjournals.org/cebp/article-pdf/28/7/1238/2285991/1238.pdf by guest on 03 November 2024

on chromosome 17q12 (*HNF1B*), variants in strong LD (pairwise r^2) with the index SNP were included in the analysis (Supplementary Table S1). To assess the aggregate contribution of these 26 GWAS loci to the estimated heritability for pancreatic cancer, 72,225 variants located within ±250 kb of the index SNP were analyzed. Together these explained 4.1% (SE = 0.8%) of the phenotypic variance for pancreatic cancer.

A total of 9,445 variants located within ±50 kb of gene boundaries (3′ UTR to 5′ UTR) of the established 11 pancreatic cancer susceptibility genes were evaluated for their contribution to the estimated heritability. Together these variants explained 0.4% (SE = 0.3%) of the phenotypic variance for pancreatic cancer.

Discussion

Our study presents a systematic investigation of the genetic architecture of pancreatic cancer. The heritability for pancreatic cancer was estimated to be 21.2% (SE = 4.8%). This estimate is substantially higher than previously reported heritability, which ranged from 9.8% to 18% (12, 20, 21). We had previously estimated the heritability of pancreatic cancer in the PanC4 GWAS to be 16.4% (95% CI, 10.4%–22.4%) applying the GREML-SC approach using 620,357 directly genotyped variants only (12). The use of imputed data in this analysis allowed greater capture of the variance explained by rare and low-frequency causal variants. In addition, GREML-LDMS approach has been shown to provide more accurate estimates than GREML-SC. GREML-LDMS allows for stratification of variants by MAF and LD, which can minimize the differences in LD between causal variants and analyzed variants and consequently reduce the bias associated with the GREML-SC estimate. Therefore, our estimate of 21.2% is a more reliable estimate of heritability than reported previously. However, it is important to note that this estimate may still not capture the full impact of very rare high-penetrance variants.

Heritability estimated using GREML or similar approaches does not fully capture variance due to rare causal variants for several reasons. Rare variants are not included in the analysis due to (i) not captured on reference panels, (ii) low imputation quality, (iii) not polymorphic before or after converting genotype probabilities to hard calls, and (iv) minor allele count below the recommended threshold of 5. In our analysis, over half (56.3%) of all imputed variants were dropped due to poor imputation quality (INFO <0.5). In addition, because GREML cannot incorporate imputation uncertainty, genotype probabilities were converted to hard calls resulting in 1.9% of imputed variants dropped due to missingness >5%. Although some of these limitations can be overcome with the use of whole genome sequencing data, the recommendation of excluding very rare variants (variants observed on 5 or fewer chromosomes) is harder to overcome and requires extremely large sample sizes. Even when the overall mutation prevalence for a given gene is >1%, which is the case for *BRCA2* (5, 8, 37) and *ATM* (5, 38) for pancreatic cancer, each mutation is only observed in 1 to 2 patients (with the exception of founder effects). This is an important limitation to consider not only when investigating the genetic architecture of pancreatic cancer but also any disease where rare high-penetrant variants are known to cause a considerable fraction of the disease.

The overall prevalence of rare high-penetrance mutations in the population analyzed is not known. However, the cases and controls included in this analysis were drawn from the same study sites reporting that 4% to 10% of patients with pancreatic cancer have rare high-penetrance mutations in established pancreatic cancer predisposition genes (4–6). The gene-based odds ratios range from 2.58 to 12.33 (6), yet the individual level variants were rare. In contrast, in the analysis we present here using GREML-LDMS, these same gene regions explain only 0.4% of the phenotypic variance for pancreatic cancer.

However, our estimates of the contribution of common variants should be more robust. Our analysis demonstrated that chromosomes 9, 7, 8, 16, 5, 2, and 1 were the top contributors to the heritability of pancreatic cancer. This is consistent with the GWAS findings as common susceptibility loci have been discovered on all these chromosomes. Because imputation captures almost all variation at common variants but only a proportion of variation at rare variants, our results when partitioned by chromosome are likely driven by common causal variants, some of which had been identified through GWAS studies.

In our analysis, known GWAS loci explained 4.1% of phenotypic variance for pancreatic cancer, leaving >10% of the common phenotypic variance unexplained. The large proportion of unexplained heritability highlights the need to continue searching for common susceptibility loci for pancreatic cancer. SNP array-based genotyping followed by imputation will remain a cost-effective strategy for gene discovery of common variants. However, larger sample sizes are needed to increase the power of current GWAS. Furthermore, as imputation reference panels of large sample size (e.g., Haplotype Reference Consortium, HRC) continue to be developed, further improvements in the power to detect associations on these variants are expected, particularly those in above average LD regions (39, 40).

Across 4 functional groups, intronic variants account for most of the phenotypic variance of pancreatic cancer (12.4%). Interestingly, 12 of 21 GWAS loci identified in the European population are mapped to intronic variants. However, it is unclear whether these variants are of direct functional significance, as opposed to simply being in LD with another functional variant in the vicinity. The coding variants, comprising about 1% of imputed variants, account for 1% of the phenotypic variance of pancreatic cancer. This is likely an underestimate since a proportion of rare or extremely rare coding variants were not imputed or were removed by quality control. It is possible that the poor imputation accuracy on rare and extremely rare variants has a greater impact on coding variants than variants in the other 3 functional groups (Supplementary Fig. S3).

Heritability of pancreatic cancer estimated in our study is still an underestimation of the overall heritability due to the imperfect characterization of genomic variation in imputation and the inherent limitations of GREML approach in capturing the contribution of very rare variants.

Disclosure of Potential Conflicts of Interest

A.L. Blackford is a consultant at the University of Maryland School of Medicine. No potential conflicts of interest were disclosed by the other authors.

Authors' Contributions

Conception and design: F. Chen, M.J. Hassan, E.A. Holly, G.M. Petersen, H.A. Risch, A.P. Klein

Development of methodology: M.J. Hassan, E.A. Holly, G.M. Petersen, A.P. Klein

Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): P. Bracci, S. Gallinger, D. Li, R.E. Neale,

Downloaded from http://aacrjournals.org/cebp/article-pdf/28/7/1238/2285991/1238.pdf by guest on 03 November 2024

S.H. Olson, G. Scelo, M. Borges, M.J. Hassan, E.A. Holly, R.J. Hung, R.C. Kurtz, I. Orlow, H. Yu, G.M. Petersen, H.A. Risch, A.P. Klein
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): F. Chen, E.J. Childs, E. Mocci, A.L. Blackford, P. Brennan, P. Duggal, M.J. Hassan, G.M. Petersen, H.A. Risch, A.P. Klein
Writing, review, and/or revision of the manuscript: F. Chen, S. Gallinger, D. Li, R.E. Neale, S.H. Olson, G. Scelo, W.R. Bamlet, A.L. Blackford, P. Brennan, K.G. Chaffee, P. Duggal, M.J. Hassan, E.A. Holly, R.J. Hung, M.G. Goggins, A.L. Oberg, I. Orlow, H. Yu, G.M. Petersen, H.A. Risch, A.P. Klein
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): G. Scelo, I. Orlow, G.M. Petersen, A.P. Klein
Study supervision: P. Brennan, M.J. Hassan, E.A. Holly, A.P. Klein

Acknowledgments

This work was supported by RO1CA154823 (PI: A.P. Klein). Genotyping Services were provided by the Center for Inherited Disease Research (CIDR). CIDR is fully funded through a federal contract from the NIH to the Johns Hopkins University; contract number HHSN268201100011I (PI: Valle awarded to A.P. Klein).

The IARC/Central Europe study was supported by a grant from the U.S. National Cancer Institute at the National Institutes of Health (R03 CA123546-02 PI: Brennan support to G. Scelo) and grants from the Ministry of Health of the Czech Republic (NR 9029-4/2006, NR9422-3, NR9998-3, MH CZ-DRO-MMCI 00209805 PI: Kollavara Support to G. Scelo).

The work at Johns Hopkins University was supported by the NCI grants P50CA062924 (PI: A. Klein), P30CA006973 (PI: Nelson support to A. Klein), and R01CA97075 (PI: Petersen support to A. Klein and M. Goggins).

The Mayo Clinic Biospecimen Resource for Pancreas Research study is supported by the Mayo Clinic SPORE in Pancreatic Cancer (P50 CA102701 PI: G. Petersen).

The Memorial Sloan Kettering Cancer Center Pancreatic Tumor Registry is supported by P30CA008748 (PI: C. Thompson support to S. Olson), the Geoffrey Beene Foundation, the Arnold and Arlene Goldstein Family Foundation, and the Society of MSKCC. The Queensland Pancreatic Cancer Study was supported by a grant from the National Health and Medical Research Council of Australia (NHMRC; grant no. 442302; PI: R. Neale). R.E. Neale is supported by a NHMRC Senior Research Fellowship (#1060183 PI: R. Neale).

The UCSF pancreas study was supported by the NIH-NCI grants (R01CA1009767 PI: E. Holly, R01CA109767-S1 PI: P. Bracci) and the Joan Rombauer Pancreatic Cancer Fund. Collection of cancer incidence data was supported by the California Department of Public Health as part of the statewide cancer reporting program; the NCI's SEER Program under contract HHSN261201000140C awarded to CPIC; and the CDC's National Program of Cancer Registries, under agreement #U58DP003862-01 awarded to the California Department of Public Health.

The Yale (CT) pancreas study is supported by NCI at the U.S. NIH, grant no. 5R01CA098870 (PI: H. Risch). The cooperation of 30 Connecticut hospitals, including Stamford Hospital, in allowing patient access, is gratefully acknowledged. The Connecticut Pancreas Cancer Study was approved by the State of Connecticut Department of Public Health Human Investigation Committee. Certain data used in that study were obtained from the Connecticut Tumor Registry in the Connecticut Department of Public Health. The authors assume full responsibility for analyses and interpretation of these data.

Assistance with genotype data quality control was provided by Cecelia Laurie and Cathy Laurie at University of Washington Genetic Analysis Center.

F. Chen was supported by the Maryland Genetics, Epidemiology, and Medicine Training Program (MD-GEM) sponsored by the Burroughs-Wellcome Fund (PI: Duggal).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received November 29, 2018; revised February 2, 2019; accepted April 18, 2019; published first April 23, 2019.

References

1. Ilic M, Ilic I. Epidemiology of pancreatic cancer. World J Gastroenterol 2016;22:9694–705.
2. American Cancer Society. Cancer facts & figures 2018 [Internet]. [cited 2018 April 25]. Available from: https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2018/cancer-facts-and-figures-2018.pdf.
3. Rahib L, Smith BD, Aizenberg R, Rosenzweig AB, Fleshman JM, Matrisian LM. Projecting cancer incidence and deaths to 2030: the unexpected burden of thyroid, liver, and pancreas cancers in the United States. Cancer Res 2014;74:2913–21.
4. Shindo K, Yu J, Suenaga M, Fesharakizadeh S, Cho C, Macgregor-Das A, et al. Deleterious germline mutations in patients with apparently sporadic pancreatic adenocarcinoma. J Clin Oncol 2017;35:3382–90.
5. Hu C, Hart SN, Polley EC, Gnanaolivu R, Shimelis H, Lee KY, et al. Association between inherited germline mutations in cancer predisposition genes and risk of pancreatic cancer. JAMA 2018;319:2401–9.
6. Yurgelun MB, Chittenden AB, Morales-Oyarvide V, Rubinson DA, Dunne RF, Kozak MM, et al. Germline cancer susceptibility gene variants, somatic second hits, and survival outcomes in patients with resected pancreatic cancer. Genet Med 2019;21:213–23.
7. Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, 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;343:78–85.
8. Zhen DB, Rabe KG, Gallinger S, Syngal S, Schwartz AG, Goggins MG, et al. BRCA1, BRCA2, PALB2, and CDKN2A mutations in familial pancreatic cancer: a PACGENE study. Genet Med 2015;17:569–77.
9. Catts ZA, Baig MK, Milewski B, Keywan C, Guarino M, Petrelli N. Statewide retrospective review of familial pancreatic cancer in delaware, and frequency of genetic mutations in pancreatic cancer kindreds. Ann Surg Oncol 2016;23:1729–35.
10. Takai E, Yachida S, Shimizu K, Furuse J, Kubo E, Ohmoto A, et al. Germline mutations in Japanese familial pancreatic cancer patients. Oncotarget 2016;7:74227–35.
11. Chaffee KG, Oberg AL, McWilliams RR, Majithia N, Allen BA, Kidd J, et al. Prevalence of germ-line mutations in cancer genes among pancreatic cancer patients with a positive family history. Genet Med 2018;20:119–27.
12. Childs EJ, Mocci E, Campa D, Bracci PM, Gallinger S, Goggins M, et al. Common variation at 2p13.3, 3q29, 7p13 and 17q25.1 associated with susceptibility to pancreatic cancer. Nat Genet 2015;47:911–6.
13. Klein AP, Wolpin BM, Risch HA, Stolzenberg-Solomon RZ, Mocci E, Zhang M, et al. Genome-wide meta-analysis identifies five new susceptibility loci for pancreatic cancer. Nat Commun 2018;9:556.
14. Amundadottir L, Kraft P, Stolzenberg-Solomon RZ, Fuchs CS, Petersen GM, Arslan AA, et al. Genome-wide association study identifies variants in the ABO locus associated with susceptibility to pancreatic cancer. Nat Genet 2009;41:986–90.
15. Petersen GM, Amundadottir L, Fuchs CS, Kraft P, Stolzenberg-Solomon RZ, Jacobs KB, et al. A genome-wide association study identifies pancreatic cancer susceptibility loci on chromosomes 13q22.1, 1q32.1 and 5p15.33. Nat Genet 2010;42:224–8.
16. Wolpin BM, Rizzato C, Kraft P, Kooperberg C, Petersen GM, Wang Z, et al. Genome-wide association study identifies multiple susceptibility loci for pancreatic cancer. Nat Genet 2014;46:994–1000.
17. Zhang M, Wang Z, Obazee O, Jia J, Childs EJ, Hoskins J, et al. Three new pancreatic cancer susceptibility signals identified on chromosomes 1q32.1, 5p15.33 and 8q24.21. Oncotarget 2016;7:66328–43.
18. Low SK, Kuchiba A, Zembutsu H, Saito A, Takahashi A, Kubo M, et al. Genome-wide association study of pancreatic cancer in Japanese population. PLoS One 2010;5:e11824.
19. Wu C, Miao X, Huang L, Che X, Jiang G, Yu D, et al. Genome-wide association study identifies five loci associated with susceptibility to pancreatic cancer in Chinese populations. Nat Genet 2011;44:62–6.
20. Lu Y, Ek WE, Whiteman D, Vaughan TL, Spurdle AB, Easton DF, et al. Most common "sporadic" cancers have a significant germline genetic component. Hum Mol Genet 2014;23:6112–8.

Downloaded from http://aacrjournals.org/cebp/article-pdf/28/7/1238/2285991/1238.pdf by guest on 03 November 2024

21. Sampson JN, Wheeler WA, Yeager M, Panagiotou O, Wang Z, Berndt SI, et al. Analysis of heritability and shared heritability based on genome-wide association studies for thirteen cancer types. J Natl Cancer Inst 2015;107: djv279.
22. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. Nature 2009; 461:747–53.
23. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, et al. Common SNPs explain a large proportion of the heritability for human height. Nat Genet 2010;42:565–9.
24. Mucci LA, Hjelmborg JB, Harris JR, Czene K, Havelick DJ, Scheike T, et al. Familial risk and heritability of cancer among twins in nordic countries. JAMA 2016;315:68–76.
25. Tenesa A, Haley CS. The heritability of human disease: estimation, uses and abuses. Nat Rev Genet 2013;14:139–49.
26. Muñoz M, Pong-Wong R, Canela-Xandri O, Rawlik K, Haley CS, Tenesa A. Evaluating the contribution of genetics and familial shared environment to common disease using the UK Biobank. Nat Genet 2016;48:980–3.
27. Evans LM, Tahmasbi R, Vrieze SI, Abecasis GR, Das S, Gazal S, et al. Comparison of methods that use whole genome data to estimate the heritability and genetic architecture of complex traits. Nat Genet 2018;50:737–45.
28. Yang J, Bakshi A, Zhu Z, Hemani G, Vinkhuyzen AAE, Lee SH, et al. Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. Nat Genet 2015;47:1114–20.
29. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nat Genet 2012;44:955–9.
30. 1000 Genomes Project Consortium, Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, et al. A map of human genome variation from population-scale sequencing. Nature 2010;467:1061–73.
31. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience 2015;4:7.
32. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 2010; 38:e164.
33. O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res 2016;44: D733–45.
34. Lee SH, Wray NR, Goddard ME, Visscher PM. Estimating missing heritability for disease from genome-wide association studies. Am J Hum Genet 2011;88:294–305.
35. National Cancer Institute. Lifetime risk (percent) of being diagnosed with cancer by site and race/ethnicity both sexes, 18 SEER areas, 2009–2011 [Internet]. [cited 2015 Nov 10]. Available from: http://seer.cancer.gov/archive/csr/1975_2011/browse_csr.php?sectionSEL=1&pageSEL=sect_01_table.15.html.
36. Yang J, Manolio TA, Pasquale LR, Boerwinkle E, Caporaso N, Cunningham JM, et al. Genome partitioning of genetic variation for complex traits using common SNPs. Nat Genet 2011;43:519–25.
37. Salo-Mullen EE, O'Reilly EM, Kelsen DP, Ashraf AM, Lowery MA, Yu KH, et al. Identification of germline genetic mutations in patients with pancreatic cancer. Cancer 2015;121:4382–8.
38. Hu C, Hart SN, Bamlet WR, Moore RM, Nandakumar K, Eckloff BW, et al. Prevalence of pathogenic mutations in cancer predisposition genes among pancreatic cancer patients. Cancer Epidemiol Biomark Prev 2016;25: 207–11.
39. Wood AR, Beaumont R, Hernandez D, Nalls M, Gibbs JR, Bandinelli S, et al. Imputation of rare variants from the new Haplotype Reference Consortium identifies associations missed by 1000 Genomes. Baltimore, MD; 2015.
40. Iglesias AI, van der Lee SJ, Bonnemaijer PWM, Höhn R, Nag A, Gharahkhani P, et al. Haplotype Reference Consortium Panel: practical implications of imputations with large reference panels. Hum Mutat 2017;38:1025–32.

Downloaded from http://aacrjournals.org/cebp/article-pdf/28/7/1238/2285991/1238.pdf by guest on 03 November 2024