Supplementary Online Content

Hu C, Hart SN, Polley EC, et al. Association Between Inherited Germline Mutations in Cancer Predisposition Genes and Risk of Pancreatic Cancer. *JAMA.* doi:10.1001/jama.2018.6228

eAppendix 1. Pancreatic Cancer Patient Research Registry

eAppendix 2. Reference Controls

eAppendix 3. Custom Sequencing of Mayo Clinic Pancreatic Cancer Cases

eAppendix 4. Bioinformatics Analysis of Sequencing Data

eAppendix 5. Cleaning and Filtering of Reference Control Sequencing Data

eAppendix 6. Statistical Analysis

eTable 1. Genes in the Custom QIAseq Panel Evaluated for Presence of Mutations in Patients With Pancreatic Cancer

eTable 2. Mutations in Individual Genes Among Cases With Pancreatic Cancer and gnomAD Controls of Different Racial and Ethnic Populations

eTable 3. Allele Count (AC) of Mutations in Mayo Clinic Cases With Pancreatic Cancer and Two Samples of Reference Controls

eTable 4. Frequency of Mutation Carriers by Gene Among Pancreatic Cancer Cases With Personal and Family History of Cancers

eTable 5. Frequency of Mutations in Individual Genes Among Cases With Pancreatic Cancer and ExAC_non_TCGA Reference Controls From Different Racial and Ethnic Populations

eTable 6. Associations Between Panel Gene Mutations and Pancreatic Cancer Using ExAC_non_TCGA Reference Controls

eTable 7. Frequency of Mutation Frequencies for Individual Genes Among Non-Hispanic White Cases With Pancreatic Cancer by Personal and Family History of Cancers

eTable 8. Comparisons of Mutation Frequencies Between Non-Hispanic White Cases With Pancreatic Cancer and gnomAD_NFE Controls

eTable 9. Associations of Panel Gene Mutations With Pancreatic Cancer Based on Non-Hispanic White Patients With Pancreatic Cancer and ExAC_NFE_non-TCGA Reference Controls

© 2018 American Medical Association. All rights reserved. **eTable 10.** Comparisons of Mutation Frequencies Between Patients With Pancreatic Cancer as the Initial Cancer and gnomAD Controls **eTable 11.** Comparisons of Mutation Frequencies Between Cases With Pancreatic Cancer With Family History of Pancreatic, Breast, Ovarian, Colorectal, or Gynecologic Cancer and gnomAD Controls

eTable 12. Comparisons of Mutation Frequencies Between Cases With Pancreatic Cancer Without Family History of Pancreatic, Breast, Ovarian, Colorectal, or Gynecologic Cancer and gnomAD Controls

eTable 13. Comparisons of Mutation Frequencies Between Cases With Pancreatic Cancer, Excluding Those With Family History of Pancreatic Cancer, and gnomAD Controls

eTable 14. Comparisons of Mutation Frequencies Between Cases With Pancreatic Cancer, Excluding Those With Family History of Breast Cancer, and gnomAD Controls

eTable 15. Comparisons of Mutation Frequencies Between Cases With Pancreatic Cancer, Excluding Those With Family History of Ovarian Cancer, and gnomAD Controls

eTable 16. Comparisons of Mutation Frequencies Between Cases With Pancreatic Cancer, Excluding Those With Family History of colorectal cancer, and gnomAD Controls

eTable 17. Comparisons of Mutation Frequencies Between Cases With Pancreatic Cancer, Excluding Those With Family History of Gynecologic Cancer, and gnomAD Controls

eTable 18. Associations of Mutations in Six Pancreatic Cancer Genes With Personal and Family History of Cancers

eTable 19. Associations Between Mutations in Each Panel Gene and Age of Diagnosis of Pancreatic Cancer Cases

eTable 20. Demographic and Clinical Characteristics of Patients With Pancreatic Cancer Included in the Survival Analysis

eFigure. Kaplan-Meier Analysis of Overall Survival for Pancreatic Cancer Cases With Mutations (Carrier) in the 6 Genes Associated With Pancreatic Cancer and Cases Without Mutations (Non-Carrier)

eReferences

This supplementary material has been provided by the authors to give readers additional information about their work.

eAppendix 1. Pancreatic Cancer Patient Research Registry

The Mayo Clinic Biospecimen Resource for Pancreas Research is an ongoing patient registry, established in 2000, under an Institutional Review Board approved protocol. We employ an "ultra-rapid" case ascertainment method to prospectively recruit research participants, which denotes identification of potential research participants at the time of first visit to Mayo Clinic, often prior to a confirmed diagnosis of pancreatic cancer. Support for the resource has been continuous from both Mayo Clinic intramural funds and the NCI-funded Mayo Clinic SPORE in Pancreatic Cancer (P50 CA102701).

All patients being evaluated for pancreatic mass or pancreatic cancer at Mayo Clinic campuses (Rochester, MN; Jacksonville, FL; or Phoenix, AZ) are eligible for participation in the Resource. Potential participants are identified by electronic search of appointment schedules and are recruited from Gastroenterology, Oncology, and Surgery clinics. Patients are usually invited in person by a study coordinator to learn about and participate in the registry protocol (~34% are approached by mail). Participation consists of completing of a consent form, risk factor and family history surveys, providing a 50 ml venous blood sample (typically 40 ml EDTA, 10 ml heparinized, and 10 ml without anti-coagulant) during the clinic visit, and authorizing study team access to medical records. The risk factor survey collected information on race and ethnicity. Race was self-identified by checking categories (American Indian/Alaskan Native; Asian/Asian American; Black/African-American, Native Hawaiian/Other Pacific Islander; White Caucasian; Multiracial (specify)). Similarly, ethnic group was self-identified by checking "Hispanic/Latino" or "Non-Hispanic/Non-Latino." Dr. Gloria Petersen, an experienced genetic epidemiologist trained in Medical Genetics and who is an American Board of Medical Genetics and Genomics -certified Ph.D. Medical Geneticist, designed the questionnaire to carefully obtain family history details in a standardized fashion. Detailed family history structure including third degree relatives was obtained and pedigrees are generated. All participant pedigrees are reviewed personally by Dr. Petersen and her study team on a semi-monthly basis to ensure that family histories obtained are consistent, with followup to the participant/family for clarification as needed. If the patient is unwilling to provide a blood sample, saliva samples are sought for genomic DNA extraction. Participants are also asked to complete follow-up surveys at 6 and 12 months. Upon returning a signed consent, participants recruited by mail complete surveys and blood sampling kits are used for biospecimen collection at their convenience in their home town or city. If components of the protocol are incomplete, reminder letters are mailed to all participants or potential participants 3 weeks and 6 weeks after the initial invitation. If there is no further response, participants are designated as non-responders.

Participant tracking, clinical data, biospecimen data, and surveys are maintained in a secure, password-protected custom Sybase database, accessible to authorized study personnel. All surveys are scanned or double entered, with built-in error checks. Medical records from clinical providers outside of Mayo Clinic are requested. Data related to health history, diagnosis, and outcomes are obtained from authorized medical records and coded into the database. Records are manually abstracted using a standardized protocol by a gastrointestinal cancer specialist/medical oncologist. Abstracted information includes type of chemotherapy and/or radiotherapy, disease staging, response to treatment and duration of therapy, and common toxicities. Records are considered as "complete" when medical record documentation about the number, type, and duration of all therapies, with documentation of cessation of therapy has been abstracted.

The study sample consisted of participants who were recruited from October 12, 2000 through March 31, 2016. In this period, we approached 6,074 patients with pancreatic adenocarcinoma. Of these, 3985 (65.6%) provided informed consent, and 3210 provided research blood specimens, including 3030 patients constituting the analytic sample. Among those who did not participate in the registry, 633 declined to participate, whereas 1448 did not respond to our invitation to be in the study (it is our experience that the majority of these participants are very ill or coping with the diagnosis of a lethal cancer, and could not provide a stated response to the research invitation); eight participants consented to be in the registry but did not provide a research blood specimen. Approximately 85.5% (2,591) of the 3,030 participants in the analytic sample were consented to

participate in the registry within 30 days of their diagnosis, with 64.7% (1,961) consented within two days. The study sample is 43.1% female, and the racial/ethnic composition is 96.6% White, 2.2% African American, 0.36% Native American, 0.4% Asian, and 0.1% Native Hawaiian/Pacific Islander; and 1.4% Hispanic ethnicity. This composition is congruent with the race and ethnicity classifications (African/African American; Hispanic, East Asian and South Asian, and Non-Hispanic White) used in the study. These classifications were used in order to match principal component derived classifications of reference controls in gnomAD. The racial/ethnic composition of the study is similar to the U.S. Census Bureau American Community Survey (2011-2015)([https://www.census.gov/programs-surveys/acs/data.html\)](https://www.census.gov/programs-surveys/acs/data.html) for the Midwest region around Mayo Clinic, which indicates that the population is 50% female; 92.6% White, 1.9% African American, 0.42% Native American, 2.5% Asian, and 0.03% Native Hawaiian/Pacific Islander. Around 4% are Hispanic ethnicity.

The biospecimen resource adheres to the principles and practices recommended by the NCI Best Practice for Biospecimen Resources [\(https://biospecimens.cancer.gov/practices/\)](https://biospecimens.cancer.gov/practices/). Mayo Clinic has highly efficient systems developed for handling biospecimens for clinical analysis, and these are adapted for research studies. Date and time stamps, and prandial status are recorded for the specimens. Blood samples drawn at Mayo Clinic are processed within 4 hours at the institutionally-supported Biospecimens Accessioning and Processing (BAP) core lab [\(http://www.mayo.edu/research/centers-programs/cancer-research/shared-resources-core-facilities-services/biospecimens-accessioning](http://www.mayo.edu/research/centers-programs/cancer-research/shared-resources-core-facilities-services/biospecimens-accessioning-processing)[processing\)](http://www.mayo.edu/research/centers-programs/cancer-research/shared-resources-core-facilities-services/biospecimens-accessioning-processing). Blood samples drawn outside of Mayo Clinic are expressed shipped directly to the BAP lab and processed, typically within 24 hours. From a 50 ml research blood sample, 4 to 5 aliquots of 1.5 ml serum, 4 to 5 aliquots of 1.5 ml plasma, two buffy coats, and genomic DNA extracted from another buffy coat sample are stored. Samples are all processed using robotic technology by standardized methods and stored rapidly in a Biostore (NEXUS) -80°C freezer and liquid nitrogen tanks (-120°C) until needed.

Genomic DNA derived from blood samples is stored for use in future research projects. Germline genetic testing is explicitly mentioned in the consent form; patients are aware this is a possible research analysis to be performed on their samples. The research registry had medical record reports of 41 participants who had received clinical genetic testing (unrelated to our research). These were presumably disclosed by a genetics professional. Among these 41, eleven had a positive genetic test result, and those results were congruent with the testing performed by the research laboratory in this report. Among the participants who had genetic analysis performed by research protocol, all were deceased, so no test results were disclosed.

A subset of pancreatic cancer cases reported in this study were previously subjected to germline genetic testing. A total of 341 patients with a family history of pancreatic cancer from the Mayo Clinic research registry, including 309 from the current study, were tested for germline pathogenic mutations in *BRCA1, BRCA2, PALB2, and CDKN2A, to assess the prevalence of mutations in these genes in the high-risk setting¹. A separate* study evaluated the prevalence of pathogenic mutations among 25 cancer susceptibility genes on the "MyRisk" panel in 302 family history positive patients, including 273 from the current study². Whole genome sequencing of germline DNA from 187 patients with a family history of pancreatic cancer, including 155 from the current study, was conducted to discover novel pancreatic cancer susceptibility genes³. In addition, 1537 randomly selected patients with pancreatic cancer that were collected from 10/12/2000-2/18/2009, including 1512 from the current study, were tested for germline pathogenic CDKN2A mutations⁴. Finally, 96 patients sequentially enrolled between 9/5/2013 and 6/17/2014, including 95 from the current study, were tested for pathogenic mutations in 22 genes as a pilot study to estimate the frequency of mutations in these genes in a sequential series not specifically selected for family history⁵. In total, only 96 sequentially recruited and 302 with a family history of pancreatic cancer were screened for a varying number of genes, leaving 2639 patients in the current study that never had a full panel of cancer susceptibility gene mutation testing.

eAppendix 2. Reference Controls

gnomAD reference controls

The Genome Aggregation Database (gnomAD) contains sequencing data with 123136 exomes and 15496 genomes from unrelated individuals sequenced as part of various disease-specific and population genetic studies. The raw sequence data was reprocessed through the same pipeline, and jointly variant-called to increase consistency across projects. The gnomAD data set contains individuals sequenced using multiple exome capture methods and sequencing chemistries. The resulting variation in coverage is incorporated into the variant frequency calculations for each variant. GnomAD was quality controlled and analyzed using the [Hail](https://hail.is/) open-source framework for scalable genetic analysis. GnomAD provides allele frequencies separately for several races and ethnic groups including non-Finnish European (NFE), which excludes Ashkenazi Jewish and Finnish European individuals; African/African American (AFR); admixed American (AMR); East Asian (EAS); and South Asian (SAS). While the gnomAD dataset overlaps with approximately 30000 common ExAC_non_TCGA controls, the substantially increased number of gnomAD controls along with updated variant calling algorithms over ExAC identified gnomAD as a valuable alternative reference control dataset.

Exome Aggregation Consortium (ExAC) reference controls

The Exome Aggregation Consortium (ExAC) contains exome sequence data from 60706 unrelated individuals sequenced as part of various disease-specific and population genetic studies. All of the raw data from these projects was reprocessed through a common pipeline. Principal component analysis (PCA) was performed to identify population clusters corresponding to individuals of European, African, South Asian, East Asian, and admixed American. Europeans were separated into individuals of Finnish and non-Finnish ancestry (NFE). ExAC also contained cancer cases from The Cancer Genome Atlas (TCGA). Exclusion of sequence data from the TCGA cases yielded ExAC non TCGA (the cancer genome atlas project) reference controls. In addition, reference controls of European non-Finnish ancestry excluding TCGA cases (ExAC_NFE_non-TCGA) were used for association studies involving non-hispanic white cases.

eAppendix 3. Custom Sequencing of Mayo Clinic Pancreatic Cancer Cases

A QIAseq custom panel of 1733 primers accounting for all coding regions and consensus splice sites from 37 cancer genes potentially involved in germline susceptibility to cancer and 126 breast cancer associated SNPs was designed for germline genetic testing of patients for inherited mutations in these predisposition genes. For this study, the focus was on results from 21 cancer predisposition genes (eTable 1) commonly found on other commercial hereditary cancer genetic testing panels for common cancer including breast, ovarian, colorectal, endometrial, and pancreatic cancer⁶⁻⁹. The QIAseq protocol was optimized for high-throughput robotic processing of DNA samples. Two pilot projects were conducted to assess the quality of multiplex PCR and sequencing using the custom panel.

The first pilot study involved 48 DNA samples extracted from blood specimens with known mutations in several genes on the panel including two large genomic rearrangements in *BRCA1*. DNA samples were amplified with the 1300 primers in two 24 sample batches using dual barcoding. Individual amplicon pools were evaluated by quantitative PCR, eGel, and Qubit analysis, pooled, and subjected to sequencing on the MiSeq. Informatics analysis was performed as described below and was blinded to mutation status of samples and mutation descriptions. All samples sequenced equivalently with on average 75% of reads mapping to target sequences. In addition, >96% of targets had >400 reads in all samples. All 48 mutations were identified including the *BRCA1* rearrangements. No false positive known or likely pathogenic mutations were identified. In a second pilot study, 48 samples from patients with pancreatic cancer, in two batches of 24, were selected based on results from prior genetic testing^{1-3,5} and were subiected to multiplex PCR using the custom QIAseq panel followed by sequencing on a HiSeq4000. >99% of targets had >20 reads in all samples. Informatics analysis was blinded to mutation content. All pre-selected mutations were identified. On the basis of these studies, the custom QIAseq panel was considered validated for analysis of DNA samples from the sequential series of patients with pancreatic cancer.

A total of 3046 genomic DNA samples extracted from peripheral blood lymphocyte samples were processed to prepare DNA libraries for Next generation sequencing. Libraries were generated using QIAseq custom panel. One sample failed library preparation. Sample libraries were sequenced in pools of 768 per lane of a HiSeq4000 (Illumina) with 150bp paired-end reads corresponding to a median coverage of 200X.

eAppendix 4. Bioinformatics Analysis of Sequencing Data

Sequencing adapters, gene-specific primers, and the QIASeq common sequencing element were removed from the sequence data as well as hard trimming the first 24bp from each read¹⁰. Reads were aligned with bwa-mem¹¹. Realignment, recalibration, Haplotype calling, Depth of Coverage, and GenotypeGVCF walkers were run from GATK v3.4-46¹².

Nucleotide reads of greater than 20X was set as the Quality Control (QC) threshold for coverage. A total of 3041 samples (99.8%) had sequencing coverage above 20X for >90% of target nucleotides. Samples were excluded if 90% of the target regions were not covered at or above 90X, due to high levels of homozygosity, or if identity by descent was suggestive of cryptic relationships.

Data analysis was performed for the resulting 3030 samples. Copy number variations were detected with PatternCNV v1.1.3¹³. Annotations were provided through the BioR toolkit¹⁴ leveraging dbNSFP v3.0¹⁵, ClinVar¹⁶, CAVA¹⁷, and population frequencies from ExAC¹⁸ (but with TCGA samples removed). Variants were viewed and filtered with VCF-Miner¹⁹. Bam files of classified pathogenic variants were viewed by IGV. All loss of function variants (nonsense, frameshift, consensus splice sites (+/-1 or 2), and any intronic or missense variants defined as pathogenic or likely pathogenic in ClinVar by two or more clinical laboratories (Ambry Genetics, SCRP, InVitae, GeneDX, Counsyl, and InSiGHT) were validated by Sanger sequencing. All suspected mosaic somatic variants (allele ratio>80:20), and truncating variants in the last 55bp of the penultimate exon or last exon that potentially avoid nonsense mediated mRNA decay and do not influence known functional domains were excluded. Similarly variants located after established cutoffs for protein function (e.g. *BRCA2* p.Tyr3208X) were excluded. Variants reported with reduced penetrance (e.g. *CHEK2* c.Ile157Thr), and variants with minor allele frequency (MAF)>0.3%, other than common founder mutations (e.g. *CHEK2* c.1100delC) were excluded.

eAppendix 5. Cleaning and Filtering of Reference Control Sequencing Data

gnomAD data cleaning and filtering:

- Restricted to gnomAD exome data
- Deleterious variant classification rules:
	- 1. Restricted to variants with AF<0.003, except known deleterious founder variants (e.g. *CHEK2* c.1100delC)
	- 2. Include LOF (loss-of-function) variants (nonsense, frameshift, +/-1,2 splice) unless Clinvar classified as benign or VUS in majority of clinical cancer genetics testing laboratories (Ambry, SCRP, InVitae, GeneDx, Emory and InSiGHT)
	- 3. Exclude missense variants and +/-≥3 splice unless Clinvar classified as pathogenic or likely pathogenic by two or more of aforementioned clinical genetics groups.
	- 4. Exclude deleterious variants with known low risk: *PMS2* c.736_741del6ins11, *TP53* p.Arg283His, 5'UTR_EX1del, p.Arg181His, p.Arg156His, CHEK2 p.Ile157Thr.
	- 5. Exclude deleterious variants not influenced by Nonsense mediated RNA decay (NMD) (Thresholds: *BRCA2* c.9924; *BARD1* c.1947, *BRIP1* c.2851, *RAD51D* c.849)
	- 6. Mark variants in *PMS2* pseudogene region (Exon9 and exon 11-15), calculate variant frequency and odds ratios without these variants.
	- 7. Review variants with Allele Count (AC) ≥15 by IGV and by frequency in control data from dbSNP.
	- 8. Stratify by populations: AFR, AMR, EAS, NFE, and SAS
- AN (allele number) was calculated as average of all variants within the coding region of a gene of interest.

ExAC data cleaning and filtering:

- Restricted to ExAC_non_TCGA exome data
- Deleterious variant classification rules:
	- o Same as in gnomAD rules 1-6
	- o Exclude ExAC non-PASS recurrent variants with allele count in ExAC>8 and tested in <20000 ExAC alleles
	- o Exclude ExAC non-PASS variants with multiple repetitive sequences called multiple times. Example: *MSH2*_c.942+2_942+6del5, *MSH2*_c.942+2_942+4delTAA, MSH2_c.942+2_942+5delTAAA, *MSH2*_c.942+2_942+3delTA, MSH2_c.942+2_942+8del7 *MSH2*_c.942+2_942+7del6.
- Stratify by populations: AFR, AMR, EAS, NFE, SAS
- AN (allele number) was calculated as average of all variants within the coding region of a gene of interest.

eAppendix 6. Statistical Analysis

Association analysis: For the overall analysis of gene specific odds ratios, the relative frequency of the race and ethnicity populations in the pancreatic cancer cases was estimated. The corresponding gnomAD reference control populations were weighted such that the relative population frequency was the same in the cases and weighted controls. Weighted logistic regression was used to estimate the odds ratio of an association between pathogenic mutations within a gene and pancreatic cancer. Confidence intervals were estimated by the profile likelihood method. All analyses were performed in R (version 4.3.2). All tests were two-sided.

For comparisons within individual populations, gene specific odds ratios for each of the genes of interest with mutations were estimated based on the combined rare, pathogenic allele count across the gene relative to the number of alleles tested (2-fold the number of individuals tested). Confidence intervals for the odds ratios were estimated based on inverting the Fisher's exact test²⁰. The method to find the shortest continuous interval consistent with the Fisher's exact test statistical significance level guarantees the p-value from the test with be in agreement with the corresponding confidence interval not including the null value within the interior of the interval. The confidence intervals were generated in R (version 4.3.2) using the *exact2x2* package²¹. To test the hypothesis that age of diagnosis of pancreatic cancer differs between mutation carriers and non-carriers, the cumulative distribution of the age of diagnosis within each subgroup and the Kolmogorov-Smirnov test was used. The test allows for the identification of any differences in the age of diagnosis distribution and was therefore considered more powerful than only testing a shift in the average age of diagnosis between carriers and non-carriers.

Associations between mutations in each gene and patient clinical characteristics were assessed using logistic regression within the pancreatic cancer cases. The odds a patient with pancreatic cancer is a mutation carrier was modeled as a function of the clinical characteristic, e.g. family history of pancreatic cancer, adjusting for the age of cancer diagnosis. The coefficient in the regression model is interpreted as an adjusted enrichment parameter, and can be used to identify subsets of patients with a higher likelihood of having a mutation in a specific gene. This is referred to as a case-case analysis since it compares pancreatic cancer cases with mutations to pancreatic cancer cases without mutations. The patient characteristics tested were personal history of any prior cancer, family history of pancreatic cancer, family history of breast cancer, and family history of colorectal cancer.

Survival analysis: The patient population for survival analysis was restricted to the subset of 2698 adenocarcinoma cases recruited at Mayo Clinic within 3 months (≤ 92 days) of an initial diagnosis. Duration of overall survival was calculated from the date of pancreatic cancer diagnosis at Mayo Clinic until the earliest of the following: death date, or last known alive date. Updates to patient vital status and the corresponding date were obtained from multiple institutional resources, and personal/family correspondence. Death information was confirmed via external vital statistics database services. Patients not known to be deceased at the time of this analysis were censored (vital status coded as not deceased) with overall survival duration calculated using the date they were last known to be alive. Date of pancreatic cancer diagnosis was defined as the date of pathological diagnosis (97%), or clinical diagnosis (3%) for those without pathology.

The primary survival analysis was a comparison of outcome for carriers versus non-carriers of pathogenic mutations in the set of 6 genes statistically significantly associated with pancreatic cancer (ATM, BRCA1, BRCA2, CDKN2A, MLH1, TP53). Copy number variants within these genes were included in the analysis. Since the number of mutation carriers within each gene was relatively small, the pre-defined analysis plan involved pooling all carriers of mutations in the genes into a single group. Median survival for mutation carriers and non-carriers was estimated using Kaplan-Meier methods, with log-rank tests used to test for significance in the survival distribution. Multivariable Cox models adjusting for age at diagnosis, gender and pancreas cancer stage (Surgically Resectable, Locally Advanced, Metastatic) were also considered, with Likelihood Ratio

tests used to test for statistical significance of the carrier status on the survival hazard. Follow-up was estimated using the Kaplan-Meier method for the censoring distribution 22 .

Abbreviations: chr, chromosome.

^a Pancreatic study: 1, genes used in pancreatic cancer analysis; 0, genes not used in analysis due to absence of mutations.

b
Information is adapted from NCBI (National Center for Biotechnology Information) Gene: http://www.ncbi.nlm.nih.gov/gene.

eTable 2. Mutations in Individual Genes Among Cases With Pancreatic Cancer and gnomAD Controls of Different Racial and

Abbreviations: AFR, African American; AMR, Hispanic; SAS, South Asian; EAS, East Asian; NFE, non-Finnish European; gnomAD, Genome Aggregation Database.

eTable 3. Allele Count (AC) of Mutations in Mayo Clinic Cases With Pancreatic Cancer and Two

┑

Abbreviations: AC, allele count; gnomAD, Genome Aggregation Database; ExAC

(ExAC_non_TCGA)Exome Aggregation Consortium_nonTCGA.

eTable 5. Frequency of Mutations in Individual Genes Among Cases With Pancreatic Cancer and ExAC_non_TCGA Reference

Abbreviations: AFR, African American; AMR, Hispanic; SAS, South Asian; EAS, East Asian; NFE, non-Finnish European; ExAC_non_TCGA,

Exome Aggregation Consortium controls excluding cancer cases from the TCGA study; TCGA, The Cancer Genome Atlas.

Abbreviations: ExAC, Exome Aggregation Consortium; ExAC_non_TCGA, Exome Aggregation Consortium controls excluding cancer cases from the TCGA study; TCGA, The Cancer Genome Atlas; OR, odds ratio.

 $^{\circ}$ Analyses do not include cases with missing (n=12) and other (n=19) race/ethnicity, leaving 2999 of 3030.

b Logistic regression weighted by race and ethnicity

 \textdegree Adjusted by Bonferroni correction for 19 genes.

eTable 7. Mutation Frequencies for Individual Genes Among Non-Hispanic White Cases With Pancreatic Cancer by Personal and Family History of Cancers

Abbreviations: gnomAD, Genome Aggregation Database; NFE, non-Finnish European; OR, odds ratio.

^a Fishers exact test.

^bAdjusted by Bonferroni correction of 19 genes.

Abbreviations: ExAC, Exome Aggregation Consortium; NFE, non-Finnish European; TCGA, The Cancer Genome Atlas; OR, odds ratio. ^a Fishers exact test.

b Adjusted by Bonferroni correction of 19 genes.

eTable 10. Comparisons of Mutation Frequencies Between Patients With Pancreatic Cancer as the Initial Cancer and

Abbreviations: gnomAD, Genome Aggregation Database; OR, odds ratio.

^a Logistic regression weighted by race and ethnicity.

 $^{\circ}$ Analyses do not include cases with missing and other race/ethnicity, leaving 2,509 of 2,535.

^c Adjusted by Bonferroni correction of 19 genes.

eTable 11. Comparisons of Mutation Frequencies Between Cases With Pancreatic Cancer With Family History of Pancreatic,

Abbreviations: gnomAD, Genome Aggregation Database; OR, odds ratio.

^a Logistic regression weighted by race and ethnicity.

b Adjusted by Bonferroni correction of 19 genes.

eTable 12. Comparisons of Mutation Frequencies Between Cases With Pancreatic Cancer Without Family History of

Abbreviations: gnomAD, Genome Aggregation Database; OR, odds ratio.

^a Logistic regression weighted by race and ethnicity.

b Adjusted by Bonferroni correction of 19 genes.

eTable 13. Comparisons of mutation Frequencies Between Cases With Pancreatic Cancer, Excluding Those With Family

Abbreviations: gnomAD, Genome Aggregation Database; OR, odds ratio.

^a Logistic regression weighted by race and ethnicity.

b Adjusted by Bonferroni correction of 19 genes.

eTable 14. Comparisons of Mutation Frequencies Between Cases With Pancreatic Cancer, Excluding Those With Family History

Abbreviations: gnomAD, Genome Aggregation Database; OR, odds ratio.

^a Logistic regression weighted by race and ethnicity.

b Adjusted by Bonferroni correction of 19 genes.

eTable 15. Comparisons of Mutation Frequencies Between Cases With Pancreatic Cancer, Excluding Those With Family History of

Abbreviations: gnomAD, Genome Aggregation Database; OR, odds ratio.

^a Logistic regression weighted by race and ethnicity.

^b Adjusted by Bonferroni correction of 19 tested genes.

eTable 16. Comparisons of Mutation Frequencies Between Cases With Pancreatic Cancer, Excluding Those With Family History of

Abbreviations: gnomAD, Genome Aggregation Database; OR, odds ratio.

^a Logistic regression weighted by race and ethnicity.

^b Adjusted by Bonferroni correction of 19 tested genes.

eTable 17. Comparisons of Mutation Frequencies Between Cases With Pancreatic Cancer, Excluding Those With Family History of

Abbreviations: gnomAD, Genome Aggregation Database; OR, odds ratio.

^a Logistic regression weighted by race and ethnicity.

^b Adjusted by Bonferroni correction of 19 tested genes.

Abbreviation: OR, odds ratio; NT, not testing because insufficient number of cases.

^a Logistic regression adjusted for age at pancreatic cancer diagnosis. *MLH1* was not included due to limited number of patients with mutations. ^b Personal history of cancer before or after pancreatic cancer. ^cAdjusted by Bonferroni correction of 5 tested genes.

eTable 19. Associations Between Mutations in Each Panel Gene and Age of Diagnosis

a
Including Multiracial, American Indian/Alaskan Native, Native Hawaiian/Other Pacific Islander.

eFigure. Kaplan-Meier Analysis of Overall Survival for pancreatic Cancer Cases With Mutations (Carrier) in the 6 Genes Associated With Pancreatic Cancer and Cases Without Mutations (Non-Carrier)

The 6 genes are: *ATM, BRCA1, BRCA2, CDKN2A, MLH1,* and *TP53*. The hazard ratio for mutation status adjusting for age at diagnosis, sex, and stage was not statistically significant (overall survival, median OS=13.6 v 11.4 months: HR=0.86, 95% CI 0.72-1.02; p=.087). The median duration of follow-up for the carrier and non-carrier groups was 101.0 (range: 0.5-154.1) months and 103.0 (range: 0.0-220.2) months, respectively. The dashed lines represent the unadjusted median survival for carriers (13.6 months) and non-carriers (11.4 months).

eReferences

- **1.** Zhen D. B., Rabe K. G., Gallinger S., et al. BRCA1, BRCA2, PALB2, and CDKN2A mutations in familial pancreatic cancer: a PACGENE study. *Genet Med.* 2015;17(7):569- 577. doi: 10.1038/gim.2014.153.
- **2.** Chaffee K. G., Oberg A. L., McWilliams R. R., et al. Prevalence of germ-line mutations in cancer genes among pancreatic cancer patients with a positive family history. *Genet Med.* 2017. doi: 10.1038/gim.2017.85.
- **3.** Roberts N. J., Norris A. L., Petersen G. M., et al. Whole Genome Sequencing Defines the Genetic Heterogeneity of Familial Pancreatic Cancer. *Cancer Discov.* 2016;6(2):166- 175. doi: 10.1158/2159-8290.CD-15-0402.
- **4.** McWilliams R. R., Wieben E. D., Rabe K. G., et al. Prevalence of CDKN2A mutations in pancreatic cancer patients: implications for genetic counseling. *Eur J Hum Genet.* 2011;19(4):472-478. doi: 10.1038/ejhg.2010.198.
- **5.** Hu C., Hart S. N., Bamlet W. R., et al. Prevalence of Pathogenic Mutations in Cancer Predisposition Genes among Pancreatic Cancer Patients. *Cancer Epidemiol Biomarkers Prev.* 2016;25(1):207-211. doi: 10.1158/1055-9965.EPI-15-0455.
- **6.** Couch F. J., Shimelis H., Hu C., et al. Associations Between Cancer Predisposition Testing Panel Genes and Breast Cancer. *JAMA Oncol.* 2017;3(9):1190-1196. doi: 10.1001/jamaoncol.2017.0424.
- **7.** Buys S. S., Sandbach J. F., Gammon A., et al. A study of over 35,000 women with breast cancer tested with a 25-gene panel of hereditary cancer genes. *Cancer.* 2017;123(10):1721-1730. doi: 10.1002/cncr.30498.
- **8.** Kurian A. W., Li Y., Hamilton A. S., et al. Gaps in Incorporating Germline Genetic Testing Into Treatment Decision-Making for Early-Stage Breast Cancer. *J Clin Oncol.* 2017;35(20):2232-2239. doi: 10.1200/JCO.2016.71.6480.
- **9.** Susswein L. R., Marshall M. L., Nusbaum R., et al. Pathogenic and likely pathogenic variant prevalence among the first 10,000 patients referred for next-generation cancer panel testing. *Genet Med.* 2016;18(8):823-832. doi: 10.1038/gim.2015.166.
- **10.** Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnetjournal.* 2011;17(1):10-12.
- **11.** Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv E-Prints.* 2013;March.
- **12.** DePristo M. A., Banks E., Poplin R., et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet.* 2011;43(5):491-498. doi: 10.1038/ng.806.
- **13.** Wang C., Evans J. M., Bhagwate A. V., et al. PatternCNV: a versatile tool for detecting copy number changes from exome sequencing data. *Bioinformatics.* 2014;30(18):2678- 2680. doi: 10.1093/bioinformatics/btu363.
- **14.** Kocher J. P., Quest D. J., Duffy P., et al. The Biological Reference Repository (BioR): a rapid and flexible system for genomics annotation. *Bioinformatics.* 2014;30(13):1920- 1922. doi: 10.1093/bioinformatics/btu137.
- **15.** Liu X., Wu C., Li C., Boerwinkle E. dbNSFP v3.0: A One-Stop Database of Functional Predictions and Annotations for Human Nonsynonymous and Splice-Site SNVs. *Hum Mutat.* 2016;37(3):235-241. doi: 10.1002/humu.22932.
- **16.** Landrum M. J., Lee J. M., Benson M., et al. ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res.* 2016;44(D1):D862-868. doi: 10.1093/nar/gkv1222.
- **17.** Munz M., Ruark E., Renwick A., et al. CSN and CAVA: variant annotation tools for rapid, robust next-generation sequencing analysis in the clinical setting. *Genome Med.* 2015;7:76. doi: 10.1186/s13073-015-0195-6.
- **18.** Lek M., Karczewski K. J., Minikel E. V., et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature.* 2016;536(7616):285-291. doi: 10.1038/nature19057.
- **19.** Hart S. N., Duffy P., Quest D. J., Hossain A., Meiners M. A., Kocher J. P. VCF-Miner: GUI-based application for mining variants and annotations stored in VCF files. *Brief Bioinform.* 2016;17(2):346-351. doi: 10.1093/bib/bbv051.
- **20.** Fay M. P. Confidence intervals that match Fisher's exact or Blaker's exact tests. *Biostatistics.* 2010;11(2):373-374. doi: 10.1093/biostatistics/kxp050.
- **21.** Fay M. P. Two-sided exact tests and matching confidence intervals for discrete data. *R Journal.* 2010;2(1):53-58.
- **22.** Korn E. L. Censoring distributions as a measure of follow-up in survival analysis. *Stat Med.* 1986;5(3):255-260.