RESEARCH BRIEF

Whole Genome Sequencing Defines the Genetic Heterogeneity of Familial Pancreatic Cancer

Nicholas J. Roberts[1,2], Alexis L. Norris[1], Gloria M. Petersen[3], Melissa L. Bondy[4], Randall Brand[5], Steven Gallinger[6], Robert C. Kurtz[7], Sara H. Olson[8], Anil K. Rustgi[9], Ann G. Schwartz[10], Elena Stoffel[11], Sapna Syngal[12], George Zogopoulos[13,14], Syed Z. Ali[1], Jennifer Axilbund[1], Kari G. Chaffee[3], Yun-Ching Chen[15], Michele L. Cote[10], Erica J. Childs[16], Christopher Douville[15], Fernando S. Goes[17], Joseph M. Herman[18], Christine Iacobuzio-Donahue[19], Melissa Kramer[20], Alvin Makohon-Moore[1], Richard W. McCombie[20], K. Wyatt McMahon[2], Noushin Niknafs[15], Jennifer Parla[20,21], Mehdi Pirooznia[17], James B. Potash[22], Andrew D. Rhim[9,23], Alyssa L. Smith[13,14], Yuxuan Wang[2], Christopher L. Wolfgang[24], Laura D. Wood[1,18], Peter P. Zandi[17], Michael Goggins[1,18,25], Rachel Karchin[15], James R. Eshleman[1,18], Nickolas Papadopoulos[2], Kenneth W. Kinzler[2], Bert Vogelstein[2], Ralph H. Hruban[1,18], and Alison P. Klein[1,16,18]

Downloaded from http://aacrjournals.org/cancerdiscovery/article-pdf/6/2/166/1824048/166.pdf by guest on 24 August 2022

ABSTRACT

Pancreatic cancer is projected to become the second leading cause of cancer-related death in the United States by 2020. A familial aggregation of pancreatic cancer has been established, but the cause of this aggregation in most families is unknown. To determine the genetic basis of susceptibility in these families, we sequenced the germline genomes of 638 patients with familial pancreatic cancer and the tumor exomes of 39 familial pancreatic adenocarcinomas. Our analyses support the role of previously identified familial pancreatic cancer susceptibility genes such as *BRCA2*, *CDKN2A*, and *ATM*, and identify novel candidate genes harboring rare, deleterious germline variants for further characterization. We also show how somatic point mutations that occur during hematopoiesis can affect the interpretation of genome-wide studies of hereditary traits. Our observations have important implications for the etiology of pancreatic cancer and for the identification of susceptibility genes in other common cancer types.

SIGNIFICANCE: The genetic basis of disease susceptibility in the majority of patients with familial pancreatic cancer is unknown. We whole genome sequenced 638 patients with familial pancreatic cancer and demonstrate that the genetic underpinning of inherited pancreatic cancer is highly heterogeneous. This has significant implications for the management of patients with familial pancreatic cancer.
Cancer Discov; 6(2); 166–75. *©2015 AACR.*

[1]Department of Pathology, Sol Goldman Pancreatic Cancer Research Center, Johns Hopkins Medical Institutions, Baltimore, Maryland. [2]Ludwig Center and the Howard Hughes Medical Institute, Sol Goldman Pancreatic Cancer Research Center, Johns Hopkins Medical Institutions, Baltimore, Maryland. [3]Department of Health Sciences Research, Mayo Clinic, Rochester, Minnesota. [4]Dan L. Duncan Cancer Center, Baylor College of Medicine, Houston, Texas. [5]Department of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania. [6]Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Ontario, Canada. [7]Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, New York. [8]Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, New York. [9]Division of Gastroenterology, Departments of Medicine and Genetics, Pancreatic Cancer Translational Center of Excellence, Abramson Cancer Center, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania. [10]Karmanos Cancer Institute, Wayne State University School of Medicine, Detroit, Michigan. [11]Department of Internal Medicine, University of Michigan, Ann Arbor, Michigan. [12]Population Sciences Division, Dana-Farber Cancer Institute, and Gastroenterology Division, Brigham and Women's Hospital, Boston, Massachusetts. [13]The Research Institute of the McGill University Health Centre, Montreal, Quebec, Canada. [14]Goodman Cancer Research Centre, McGill University, Montreal, Quebec, Canada. [15]Department of Biomedical Engineering, Institute for Computational Medicine, Johns Hopkins University, Baltimore, Maryland. [16]Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland. [17]Department of Psychiatry and Behavioral Sciences, Johns Hopkins Medical Institutions, Baltimore, Maryland. [18]Department of Oncology, Sol Goldman Pancreatic Cancer Research Center, Johns Hopkins Medical Institutions, Baltimore, Maryland. [19]Memorial Sloan Kettering Cancer Center, New York, New York. [20]Stanley Institute for Cognitive Genomics, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York. [21]inGenious Targeting Laboratory, Ronkonkoma, New York. [22]Department of Psychiatry, University of Iowa, Iowa City, Iowa. [23]Department of Medicine, University of Michigan, Ann Arbor, Michigan. [24]Department of Surgery, Sol Goldman Pancreatic Cancer Research Center, Johns Hopkins Medical Institutions, Baltimore, Maryland. [25]Department of Medicine, Sol Goldman Pancreatic Cancer Research Center, Johns Hopkins Medical Institutions, Baltimore, Maryland.

doi: 10.1158/2159-8290.CD-15-0402

©2015 American Association for Cancer Research.

INTRODUCTION

Pancreatic ductal adenocarcinoma (PDAC) is a devastating disease, with a reported 5-year survival rate of 7% (1). Over 48,000 PDACs are estimated to have been diagnosed in the United States in 2015. Of these, up to 10% occur in families with at least two affected first-degree relatives, and these are designated familial pancreatic cancers (FPC; ref. 2). Individuals with a family history of PDAC carry a 2.3- to 32-fold increased risk of developing the disease, depending upon the number of affected family members (3). In some FPC kindreds, the aggregation of pancreatic cancer may be due to environmental factors or stochastic events, but many are thought to be caused by inherited genetic susceptibility (4).

Knowledge of the genes responsible for an inherited susceptibility to pancreatic cancer is important for a number of reasons. First, early detection can be targeted to mutation carriers, and pancreatic neoplasms detected at an earlier stage, when therapeutic interventions with curative potential are still available (5). Second, as most previously reported FPC susceptibility genes also increase risk for malignancies other than pancreatic cancer, these extra-pancreatic neoplasms can be screened for as well (6). Third, elucidation of the genetic basis of FPC susceptibility offers opportunities for personalized therapies, as demonstrated by patients whose pancreatic cancers harbor defects in homologous recombination arising from biallelic inactivation of *BRCA1*, *BRCA2*, or *PALB2*. In these patients, targeting DNA repair with poly(ADP-ribose) polymerase 1 (PARP-1) inhibitors, platinum compounds, or mitomycin C can result in major therapeutic benefits (7). Finally, identifying causal FPC genes will provide novel insights into PDAC tumorigenesis.

Recent advances in sequencing technology provide an unbiased way to search for the genes underlying disease susceptibility (8). Using this approach, *PALB2* and *ATM* were identified as FPC susceptibility genes, together explaining 3% to 5% of FPC cases (8, 9). In a further 8% to 15% of patients with FPC, the increased risk of pancreatic cancer can be attributed to 10 other previously reported FPC susceptibility genes, including *BRCA1*, *BRCA2*, *CDKN2A*, *MLH1*, *MSH2*, *MSH6*, *PMS2*, *PRSS1*, *STK11*, and *TP53* (10–16). The genetic basis underlying disease susceptibility in the remaining 80% to 90% of patients with FPC is unknown.

To explore the genetic basis of FPC in detail and identify candidate susceptibility genes, we performed whole genome sequencing on the germline DNA of 638 patients with FPC from 593 kindreds. This sequencing was supplemented with the whole exome sequencing of surgically resected PDACs from 39 patients with FPC. The results identify novel candidate FPC susceptibility genes and validate the importance of established FPC genes. In addition, our results suggest that somatic mutations in hematologic malignancy driver genes can confound the findings of germline genomic sequencing studies in older populations. Finally, we provide an unprecedentedly large resource of deep, whole genome sequencing data that can be used for pancreatic cancer research.

Note: Supplementary data for this article are available at Cancer Discovery Online (http://cancerdiscovery.aacrjournals.org/).

Corresponding Authors: Bert Vogelstein, Ludwig Center, Johns Hopkins Medical Institutions, 1650 Orleans Street, Baltimore, MD 21231. Phone: 410-955-8878; Fax: 410-955-0548; E-mail: vogelbe@jhmi.edu; Nicholas J. Roberts, Department of Pathology, Johns Hopkins Medical Institutions, 1550 Orleans Street, Baltimore, MD 21231. Phone: 410-502-5162; E-mail: nrobert8@jhmi.edu; Kenneth W. Kinzler, E-mail: kinzlke@jhmi.edu; Ralph H. Hruban, Department of Pathology, Johns Hopkins Medical Institutions, Baltimore, 600 North Wolfe Street, MD 21287. Phone: 410-955-9791; E-mail: rhruban@jhmi.edu; and Alison P. Klein, Department of Oncology, Johns Hopkins Medical Institutions, 1550 Orleans Street, Baltimore, MD 21231. Phone: 410-955-3512; E-mail: aklein1@jhmi.edu

RESULTS

Sample Selection and Sequencing

A total of 638 patients with FPC (Table 1) were selected from 10 registries across North America. Patients with FPC known to have a deleterious variant in a previously reported FPC susceptibility gene were excluded from the study to maximize the opportunity to discover novel susceptibility genes. Whole genome sequencing generated an average of 135.6 Gb of data per patient (range, 102.2–253.8 Gb), resulting in an average coverage of 39.8-fold (range, 29.8–71.1) per genome, with 98.2% (range, 97.9%–98.6%) and 96.0% (range, 92.8%–97.2%) of bases covered at least 1 and 10 times, respectively. An average of 3,742,720 single-nucleotide variants (SNV) were identified per patient (range, 3,623,824–4,554,474) with 93.4% (range, 86.9%–94.1%) of variants present in the database of single-nucleotide polymorphisms (dbSNP; ref. 17). The integrity of our pipeline for calling sequence variants was supported by the excellent agreement between whole genome sequencing and Illumina HumanOmni2.5 SNV array (99.2%;

Table 1. Characteristics of the whole genome sequenced patients with FPC

Classification	Number
Cohort	
FPC patients	638
FPC kindred	593
Age, y	
Less than 50	35
50–59	124
60–69	214
70–79	185
80+	73
Unknown	7
Genetic ancestry	
African	18
Asian	8
Caucasian	612
Affected relatives	
2	358
3	196
4 or more	84
DNA origin	
Blood	454
Lymphoblastoid cell line	158
Tissue	26

Downloaded from http://aacrjournals.org/cancerdiscovery/article-pdf/6/2/166/1824048/166.pdf by guest on 24 August 2022

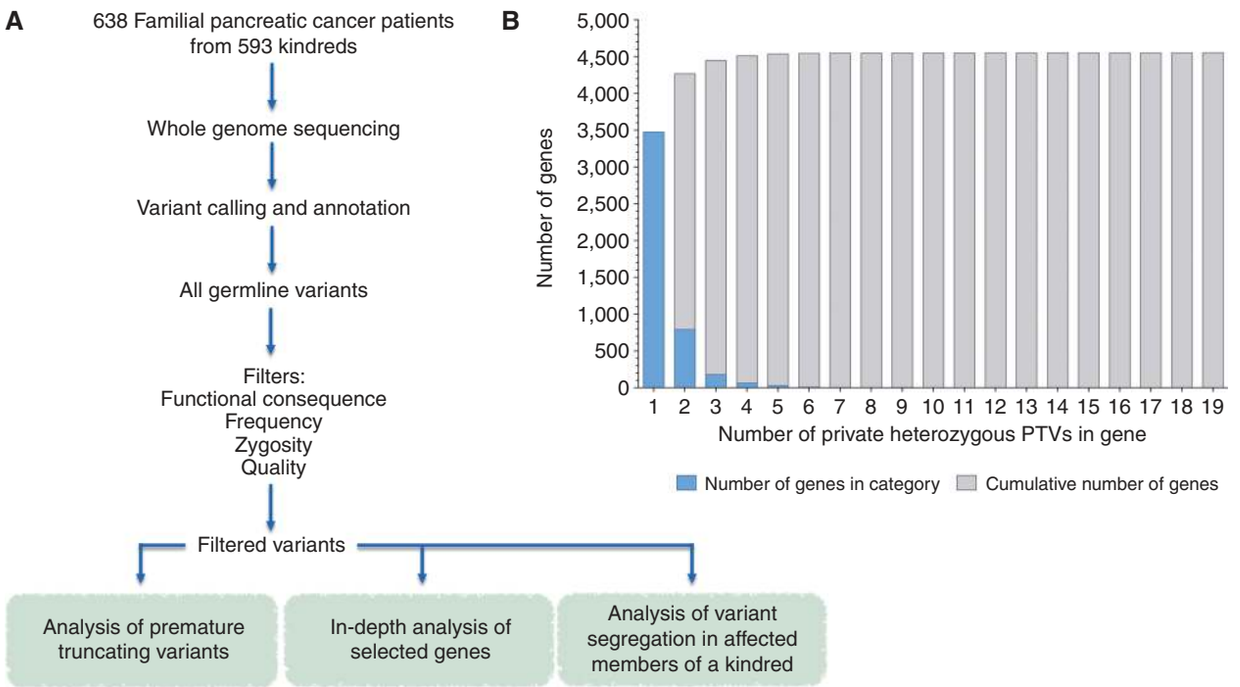

Figure 1. **A,** overview of filter-based strategy to identify novel candidate FPC susceptibility genes. 638 patients with FPC from 593 kindreds. Patients with FPC were selected from 10 high-risk family registries in North America with diverse ascertainment screens, including internet recruitments, medical genetics clinics, and tertiary care facilities. Demographic and sample data for the 638 patients with FPC are shown in Table 1. FPC patient samples were whole genome sequenced and aligned to the human genome build hg19 before variant calling and annotation. One patient with FPC from each kindred was arbitrarily selected for filter-based analyses. All germline variants in selected patients with FPC were identified and, depending on analysis, filtered by (i) functional consequence of variant, (ii) frequency of the variant in patients with FPC and variant databases (the 1000 Genomes Project and Exome Variant Server), (iii) zygosity of variant, and (iv) variant quality. The following analyses were then performed on filtered variants: (i) analysis of premature truncating variants (PTV), (ii) in-depth analysis of selected genes, and (iii) analysis of variant segregation in affected members of a kindred. **B,** distribution of private heterozygous PTVs in patients with FPC selected for filter-based analyses. Blue bars, number of genes within each PTV category. Gray bars, cumulative number of genes.

range, 99.0%–99.3%). There were an average of 328,689 (range, 279,767–399,378) insertions and 343,418 (range, 305,159–421,483) deletions per patient. The insertions averaged 23 bp (range, 1–300 bp) and the deletions 11 bp (range, 1–300 bp). The genetic ancestry of patients with FPC was determined using Local Ancestry in adMixed Populations (LAMP). Patients with FPC were predominantly of European ancestry (95.9%), but patients of African (2.8%) and Asian (1.3%) ancestry were also represented (Table 1). Identity-by-descent analysis confirmed expected familial relationships.

Analysis of Premature Truncating Variants

Given that most high-penetrance disease-associated variants so far identified are located in coding regions (18), we focused our analyses on genetic variants located in these regions. The functional significance of missense variants is often unclear. We therefore began our analysis with premature truncating variants (PTV), as these almost always affect protein function. As FPC is a rare disease and common PTVs are less likely confer a high risk of FPC susceptibility due to negative selective pressures, we concentrated our analyses on private heterozygous PTVs. We arbitrarily selected one sequenced member from each of the 593 FPC kindreds and positively filtered variants using the following criteria (Fig. 1A): (i) nonsense variants, splice-site variants, and frameshift INDELs; (ii) heterozygous in the germline; (iii) less than 0.5% minor allele frequency (MAF) in the 1000 Genomes Project or Exome Variant Server (EVS); and (iv) present in only one patient with FPC, i.e., "private" (19, 20). Finally, we selected high-quality rare heterozygous PTVs by filtering for variants with (i) a mappability score of at least 0.5 and (ii) no more than one additional genomic locus as assessed by BLAT (21, 22). Using these filters, we identified 6,114 private heterozygous PTVs, in 4,553 genes.

In order to identify novel FPC susceptibility genes, we then ranked 20,049 coding genes by the number of private heterozygous PTVs that they harbored (Supplementary Table S1). Several of the 12 previously reported FPC susceptibility genes were highly ranked, providing support for this general approach. For example, the highest ranked gene was *ATM*, with 19 private heterozygous PTVs. Similarly, *PALB2* (5 heterozygous PTVs) and *CDKN2A* (4 heterozygous PTVs) were also ranked highly. Although most genes harbored only one private heterozygous PTV and presumably do not play a common role in FPC susceptibility, 1,077 genes contained 2 or more private heterozygous PTVs (Fig. 1B). In particular, 16 genes previously identified as an FPC susceptibility gene, cancer driver gene, or DNA repair gene contained 3 or more

Downloaded from http://aacrjournals.org/cancerdiscovery/article-pdf/6/2/166/1824048/166.pdf by guest on 24 August 2022

Table 2. Genes with three or more private heterozygous PTVs in unrelated patients with FPC

Gene	Number of heterozygous PTVs in FPC kindred	Comment
ATM	19	FPC susceptibility gene; cancer driver gene; DNA repair gene
TET2	9	Cancer driver gene
DNMT3A	7	Cancer driver gene
POLN	6	DNA repair gene
POLQ	6	DNA repair gene
ASXL1	5	Cancer driver gene
BRCA2	5	Cancer driver gene; DNA repair gene
PALB2	5	FPC susceptibility gene; DNA repair gene
CDKN2A	4	FPC susceptibility gene; cancer driver gene; DNA repair gene
FANCG	4	DNA repair gene
BUB1B	3	DNA repair gene
ESCO2	3	DNA repair gene
FANCC	3	DNA repair gene
FANCM	3	DNA repair gene
MSH4	3	DNA repair gene
RAD54L	3	DNA repair gene

private heterozygous PTVs and represent the most promising candidates for further study (Table 2; refs. 23, 24).

We detected private heterozygous PTVs in *TET2* ($n = 9$), *DNMT3A* ($n = 7$), and *ASXL1* ($n = 5$; Table 2). Recent evidence has indicated that somatic mutations in genes contributing to hematologic malignancies are detectable in the blood of older individuals, suggesting a potentially preleukemic clonal hematopoiesis (25–28). As DNA used for whole genome sequencing was primarily derived from peripheral white blood cells (Table 1), when possible, we sequenced these mutations in DNA from a second non-blood source (two patients, FPC0072 and FPC0083, in Supplementary Table S2). In both cases, the mutation was not found or was found at much lower levels than observed in DNA from blood, suggesting these mutations may be somatic in nature.

It is possible that rare heterozygous PTVs in our FPC cohort contribute to susceptibility, as would be the case for founder mutations. Allowing the same heterozygous PTV to occur in as many as 10 patients with FPC (rather than in only one patient with FPC) did not significantly change the outcome our analysis. Specifically, using the same patients with FPC, 9,689 heterozygous PTVs across 5,116 genes were observed, and 80% of these were also identified when filtering for only private mutations.

In-Depth Analysis of Selected Genes

We conducted an in-depth analysis of 87 genes that included (i) previously reported FPC susceptibility genes, (ii) genes associated with hereditary cancers, and (iii) genes mutated in hereditary pancreatitis (Supplementary Table S3). As these genes had already been associated with disease, we were able expand our filter beyond just PTVs to evaluate all variants based on their functional consequences, minor allele frequencies in the 1000 Genomes Project and EVS, and ClinVar classification (19, 20, 29).

We identified SNVs, insertions and deletions (INDEL) less than 300 bp in length, and structural variant deletions (SVD) greater than 300 bp in length that affected the coding regions of these 87 genes. Variants were classified as either benign, of unknown significance (VUS), or deleterious according to the criteria detailed in Table 3. Among all 638 patients with FPC sequenced, 92,933 sequence variants were identified in these 87 genes (Supplementary Table S4). Among the 593 unrelated patients with FPC, 86,486 sequence variants were identified, 194 of which were defined as deleterious. In the 12 reported FPC susceptibility genes, there were 62 deleterious variants in 58 FPC kindreds (9.8% of FPC kindreds; 95% confidence interval: 7.6%–12.4%). In 32 patients with FPC, deleterious variants in two or more of the 87 genes analyzed in depth were observed (Supplementary Table S5). Of these patients, four had deleterious variants in two FPC susceptibility genes: 1 patient had an *ATM* and a *PALB2* deleterious variant, 1 patient had two deleterious *TP53* variants, and 2 patients had deleterious variants in both *BRCA1* and *BRCA2*. A further 17 patients had a deleterious variant in an FPC susceptibility gene in addition to a deleterious variant in a hereditary cancer or hereditary pancreatitis gene.

It should be noted that patients with FPC known to have a deleterious variant in reported FPC susceptibility genes prior to the start of this study were not selected for sequencing. Therefore, our analysis underestimates the true prevalence of previously reported FPC susceptibility genes for which clinical testing is not uncommon, such as *BRCA1*, *BRCA2*, *CDKN2A*, and *PALB2*. At the time of patient selection, *ATM*

Downloaded from http://aacrjournals.org/cancerdiscovery/article-pdf/6/2/166/1824048/166.pdf by guest on 24 August 2022

Table 3. Classification of germline variants

Variant type	MAF	ClinVar
Benign		
Any	>0.5%	—
Synonymous SNV	≤0.5%	—
VUS		
Missense SNV	≤0.5%	Not pathogenic or probable—pathogenic
In-frame INDEL	≤0.5%	Not pathogenic or probable—pathogenic
Deleterious		
Frameshift INDEL	≤0.5%	—
Nonsense SNV	≤0.5%	—
Splicing SNV or INDEL	≤0.5%	—
Missense SNV	≤0.5%	Pathogenic or probable—pathogenic
In-frame INDEL	≤0.5%	Pathogenic or probable—pathogenic
SV deletion	—	—

was not commonly tested, and we identified 21 patients from 20 kindreds with deleterious variants in this gene (3.4% of FPC kindreds; 95% confidence interval: 2.2%–5.2%).

In addition to publicly available data from the 1000 Genomes Project and EVS, we compared our findings in the 87 selected genes to whole exome sequencing data from 967 unrelated participants of European ancestry from the Bipolar Case–Control Study (BCCS; ref. 30). In BCCS samples, call rates across the 87 genes averaged 0.889. Structural variant data were not available from the BCCS; therefore, analysis was limited to SNVs and INDELs (Supplementary Tables S4 and S6). First, we compared deleterious variants in the 593 FPC kindreds to BCCS samples. Five genes were associated with FPC at a point-wise level of 0.05: *ATM* [$P = 1.2 \times 10^{-7}$; Benjamini–Hochberg (q) value = 1.1×10^{-5}], *CDKN2A* ($P = 8.8 \times 10^{-6}$; $q = 4.0 \times 10^{-4}$), *APC* ($P = 0.0174$; $q = 0.3786$), *PALB2* ($P = 0.0079$; $q = 0.2290$), and *BRCA1* ($P = 0.0317$; $q = 0.5523$). In addition, five genes had P values between 0.05 and 0.10: *BUB1B* ($P = 0.0548$; $q = 0.6306$), *FANCC* ($P = 0.0548$; $q = 0.6306$), *BRCA2* ($P = 0.0671$; $q = 0.6306$), *CPA1* ($P = 0.0725$; $q = 0.6306$), and *FANCG* ($P = 0.0725$; $q = 0.6306$). We then limited the analysis to 245 unrelated patients with FPC from kindreds with three or more affected relatives (Supplementary Table S7). Five genes had a significant difference in the number of deleterious mutations at a point-wise level of 0.05: *ATM* ($P = 4.4 \times 10^{-6}$; $q = 4.0 \times 10^{-4}$), *CDKN2A* ($P = 1.3 \times 10^{-5}$; $q = 5.0 \times 10^{-4}$), *APC* ($P = 0.0013$; $q = 0.0376$), *BUB1B* ($P = 0.0082$; $q = 0.1430$), and *PALB2* ($P = 0.0082$; $q = 0.1430$; Supplementary Table S8). These associations remained significant when the analysis was restricted to individuals with greater than 80% European genetic ancestry (Supplementary Table S8).

Analysis of Variant Segregation in Affected Members of a Kindred

We hypothesized that a deleterious variant shared among family members with pancreatic cancer was more likely to be associated with pancreatic cancer susceptibility. Therefore, we assessed segregation of: (i) private heterozygous PTVs across the exome; and (ii) deleterious variants identified from in-depth analysis of hereditary cancer genes, in 38 FPC kindreds (83 patients with FPC), where DNA from more than one affected family member was sequenced.

We identified 904 private heterozygous PTVs in the patients with FPC of the 38 kindreds. Of these, 112 private heterozygous PTVs, in 110 genes, were present in all sequenced affected family members of a kindred and therefore segregated with PDAC (Supplementary Table S9). Most of these genes (70 of 110; 63.6%) were found to have private heterozygous PTVs in only a single FPC kindred. Of note, 5 of the 110 genes were previously associated with DNA repair or are cancer driver genes: *ATM*, *CDKN2A*, *NUDT1*, *POLD1*, and *RECQL*. However, only *ATM* and *CDKN2A* were found to have private heterozygous PTVs in more than one FPC kindred (23, 24).

Seventeen deleterious variants in one of the 87 genes analyzed in-depth occurred in a patient with FPC from a family in which another affected family member had been sequenced. Deleterious variants included six frameshift deletions, two nonsense SNVs, two splice-site SNVs, and nine nonsynonymous SNVs (Supplementary Table S10). In 13 of the 17 cases, the deleterious variant did *not* perfectly segregate among affected family members. For example, we observed nonsegregation of deleterious variants in *ATM* (one kindred with 1 of 2 affected members carrying a variant), *CDKN2A* (two kindreds each with 2 of 3 affected members carrying a variant), *BRCA1* (one kindred with 1 of 2 members carrying a variant), and *PALB2* (one kindred with 1 of 2 members carrying a variant).

Somatic Alterations in FPCs

Hereditary cancer susceptibility genes are often tumor suppressors in which a deleterious variant in the germline of an individual is accompanied by a second somatic event resulting in biallelic loss of the gene in the tumor (18, 31). To help identify candidate susceptibility genes through the identification of such second somatic "hits," we sequenced the exomes of 39 pancreatic cancers resected from patients with FPC. Whole exome sequencing rather than whole genome

Downloaded from http://aacrjournals.org/cancerdiscovery/article-pdf/6/2/166/1824048/166.pdf by guest on 24 August 2022

sequencing was conducted because PDACs often contain a significant proportion of nonneoplastic cells (even after careful microdissection). Therefore, we could increase coverage to 100 times, enhancing sensitivity of somatic mutation detection. Because of the low neoplastic content of these lesions, we did not identify losses of heterozygosity or changes in copy number, and examined only somatic mutations. Exome sequencing revealed 1,409 somatic mutations, with an average of 36 mutations per tumor (Supplementary Table S11). As expected, somatic mutations in *KRAS* and *TP53* were the most common, occurring in 84.6% and 71.8%, respectively (Supplementary Table S12; ref. 32). Other genes somatically mutated in the cancers included *SMAD4* (33.3%) and *CDKN2A* (12.8%). The prevalence of *KRAS*, *TP53*, *SMAD4*, and *CDKN2A* mutations is similar to previous reports of both sporadic and familial pancreatic cancer (7, 32–34). Hereditary cancer genes were somatically mutated in the 39 PDACs, including *FANCM* in two tumors, and *BRCA2*, *BUB1B*, *CREBBP*, *FLCN*, *PTCH1*, *PTEN*, *RB1*, *TSC2*, and *WAS* in one tumor each (Supplementary Table S12). Patients with FPC with a somatic mutation in one of these genes did not have a deleterious germline variant in the same gene. Furthermore, one patient had a deleterious germline variant in a previously reported FPC susceptibility gene (FPC0347; *PALB2*) but did not have a second somatic mutation in the tumor. However, loss of heterozygosity at this locus could not be ruled out.

Of the 4,553 genes that harbored at least one private heterozygous PTV in our genome-wide analysis, 366 (8.0%) were also found to have a somatic mutation in at least one sequenced pancreatic tumor (Supplementary Table S1). Of these 366 genes, 113 had multiple private heterozygous PTVs and for 74 there were more private heterozygous PTVs in the FPC kindreds than similarly analyzed BCCS samples. Of note, 5 of these 74 genes, *BUB1B*, *CDKN2A*, *RAD54L*, *RFC1*, and *TP53*, are associated with DNA repair or known to be a cancer driver gene (Supplementary Table S1; refs. 23, 24).

DISCUSSION

The genetic basis of FPC is poorly defined. We conducted germline whole genome sequencing of 638 patients with FPC and demonstrate that inherited pancreatic cancer is highly heterogeneous. This heterogeneity has significant implications for the management of patients with a family history of this disease.

Our results provide strong evidence in support of previously reported FPC susceptibility genes, such as *ATM*, *BRCA2*, *CDKN2A*, and *PALB2*, elevating risk of pancreatic cancer. As well, our study suggests that deleterious variants in the candidate genes *BUB1B*, *CPA1*, *FANCC*, and *FANCG* are more frequent in patients with FPC (Table 2 and Supplementary Table S8). Interestingly, many of these candidate genes are involved in processes regulating DNA repair or chromosomal stability, just as are the previously identified *ATM*, *BRCA2*, and *PALB2* genes.

BUB1B encodes a protein involved in spindle-assembly checkpoint, and germline mutations in *BUB1B* are known to predispose to premature chromatid separation syndrome and other cancer types (4, 35). Heterozygous, inactivating mutations in *BUB1B* were present in three patients with FPC. In one patient with FPC, a second affected relative was available, and in this case, the *BUB1B* variant was not present (Supplementary Table S10). Still, incomplete segregation of FPC susceptibility genes such as *ATM*, *BRCA1*, *CDKN2A*, and *PALB2* is not uncommon in FPC kindreds (Supplementary Table S10) and in comparison samples from BCCS samples, no deleterious variants in *BUB1B* were identified. Additional support for *BUB1B* as a candidate pancreatic cancer susceptibility gene can be found in variant databases such as EVS and ExAC, where the sum of minor allele frequencies of *BUB1B* PTVs in the general population is 0.00024 and 0.00082 respectively. These frequencies are below the level observed in all FPC kindreds (0.00253) and the most severely affected FPC kindreds with 3 or more affected members (0.00612). Our observation of a somatic *BUB1B* mutation in one of the 39 pancreatic cancers sequenced provides further evidence in support of *BUB1B* as a candidate susceptibility gene.

Our results also suggest deleterious germline variants in *CPA1* may be more frequent in patients with FPC. Four heterozygous nonsense variants in *CPA1* were found in patients with FPC (3 chr7:130020952_C>T; p.R27X variants and 1 chr7:130021680_C>A; p.Y119X variant). This finding is intriguing given that deleterious variants in this gene have recently been shown to predispose to chronic pancreatitis and that chronic pancreatitis is strongly associated with an increased risk of pancreatic cancer (16, 36). Two of the patients with FPC with a deleterious *CPA1* variant reported a history of pancreatitis approximately 1 year before diagnosis. Intriguingly, 3.1% of recently diagnosed patients with pancreatic cancer report a history of pancreatitis within a year of diagnosis (37). As the p.R27X variant identified in patients with FPC has previously been shown to be functionally defective, a history of subclinical chronic pancreatitis cannot be ruled out (36).

Ten patients harbored the same deleterious variant in *APC* (chr5:112175211_T>A; p.I1307K). As this *APC* variant is prevalent in Jewish populations and the proportions of patients with FPC and BCCS samples of Jewish ancestry are unknown, further studies to validate this association are warranted, especially considering the equivocal role that the *APC* gene broadly plays in FPC susceptibility (38–40).

There are at least three observations from our study that are likely to have an impact on research involving other hereditary cancers. The first is that FPC appears to be heterogeneous with respect to its genetic underpinnings. Although this statement is not surprising given prior research on FPC, it was possible that a previously undiscovered gene was responsible for the majority of FPC cases. Our data, obtained from a very large number of FPC cases, largely exclude this possibility at least for truncating mutations within the coding regions of the 20,049 recognized protein-coding genes.

Second, and more subtly, we observed that variants in well-recognized FPC susceptibility genes were often not present in other affected individuals from the same family (Supplementary Table S10). Segregation of variants among affected members is the hallmark of susceptibility to any disease and provides the conceptual foundation for linkage analysis. The extent of phenocopies in our study, though surprising in some respects, is not without precedent. In one of the first

Downloaded from http://aacrjournals.org/cancerdiscovery/article-pdf/6/2/166/1824048/166.pdf by guest on 24 August 2022

reports of a gene conferring susceptibility to cancer, it was noted that a particular *TP53* gene mutation was not present in a young patient with breast cancer from a Li–Fraumeni family (41). This patient, in retrospect, was obviously a phenocopy. In our cases, the lack of coinheritance could be explained by either phenocopies, the possibility that multiple deleterious variants are responsible for the phenotype within that family, or the possibility that the variant we classified as deleterious was not responsible for the phenotype. It is of interest to note that many of the previously reported pancreatic cancer susceptibility genes, such as *BRCA1*, *PALB2*, *CDKN2A*, and *ATM*, as well as our candidate genes, would be eliminated as susceptibility genes if phenocopies were not considered (Supplementary Table S10).

A third point raised by our results pertains to the nature of the peripheral white blood cell DNA that is used for virtually all large-scale genomic studies. We observed multiple private heterozygous PTVs in *ASXL1*, *DNMT3A*, and *TET2* in patients with FPC, which would suggest that they are FPC susceptibility genes. Given these genes have been shown to be somatically mutated in the blood of phenotypically normal individuals (25–28), we attempted to confirm the germline origin of these variants in 2 patients with FPC by sequencing DNA from a second tissue. In both cases, our results indicated that these variants were somatic in nature. These unexpected observations emphasize that DNA derived from peripheral white blood cells cannot always be equated with germline DNA, especially in older individuals. This is of particular importance given that many sequencing studies, including ours, use publicly available control data where age data are not available and the age distribution of the controls may be different from that of the study population. Thus, somatic mutations in peripheral white blood cells could lead to false-positive associations, particularly for diseases strongly related to aging.

Finally, we focused on rare PTVs because these variants alter their encoded proteins in an extreme fashion and are predicted to inactivate them. There are, however, other types of variants that may contribute to FPC susceptibility. Further studies will be necessary to delineate the role of missense and noncoding variants in FPC, as current algorithms to discriminate deleterious from benign variants are not accurate. In addition, large INDELs may be poorly detected by our sequencing method. As such, alternative approaches may be necessary to determine the contribution of large INDELs to FPC susceptibility. Recognizing the need for long-term research, we chose to use whole genome rather than exome sequencing. Whole genome analyses provide a more complete resource to the pancreatic cancer research community. As more information about gene regulatory regions becomes available through projects such as ENCODE, and as more control individuals' whole genome sequences becomes publicly available, the utility of the resource provided herein will correspondingly increase.

METHODS

Institutional Approval and Informed Consent

This study followed the recommendations of the Declaration of Helsinki. Each study site obtained Institutional Review Board approval for their study protocols. Informed consent was obtained from all study participants at their respective institution.

FPC Patient Samples

Patients with FPC were enrolled into the National Familial Pancreatic Tumor Registry (NFPTR) at Johns Hopkins or one of the FPC registries participating at the following sites: Dana-Farber Cancer Institute, Karmanos Cancer Institute, Mayo Clinic, McGill University Health Centre, Memorial Sloan Kettering Cancer Center, Mount Sinai Hospital, University of Michigan, University of Pennsylvania, and University of Pittsburgh. FPC families are defined as kindreds with at least one pair of first-degree relatives diagnosed with PDAC. When possible, all cancer diagnoses in each kindred were pathologically confirmed. Patients with a previously known deleterious variant in a previously reported FPC susceptibility gene (*ATM*, *BRCA1*, *BRCA2*, *CDKN2A*, *MLH1*, *MSH2*, *MSH6*, *PALB2*, *PMS2*, *PRSS1*, *STK11*, and *TP53*) were excluded from the study. Germline DNA samples were obtained from either blood, Epstein–Barr virus transformed peripheral blood lymphocytes [lymphoblastoid cell line (LCL)], or nontumor tissue.

Whole Genome Sequencing of Germline FPC Patient Samples

A total of 638 FPC patient samples were whole genome sequenced and genotyped with the HumanOmni2.5-8v1 array (Illumina) by Personal Genome Diagnostics. Briefly, 3 μg of genomic DNA per patient sample was sequenced using the Illumina Whole Genome Sequencing Service with the Illumina HiSeq 2000 (Illumina), generating 200 base pair (bp; 2 × 100 bp reads) per fragment in the final library. Sequence reads were analyzed and aligned to the human reference genome (hg19) using Illumina CASAVA v1.7 and ELAND v.2 software (Illumina). Variants were annotated using CRAVAT with (i) functional consequence in RefSeq gene transcripts, (ii) zygosity, (iii) MAF using publicly available variant databases (1000 Genomes Project and Exome Variant Server), and (iv) presence in ClinVar (19, 20, 29, 42, 43). For each variant, mappability score for a 100-bp read as well as the number of BLAT genomic locations for a 101-bp sequence centered about the variant mapping to 80 to 120 bp of the reference genome with at least 90% identity were determined (21, 22).

Identity by Descent and Local Ancestry in adMixed Populations Analysis of Patients with FPC

Identity-by-descent (IBD) sharing analysis was performed on patients with FPC using 22,458 independent SNPs with an R^2 cutoff of 0.0001 and outside regions of high linkage disequilibrium (LD). Reported familial relationships were confirmed.

Local Ancestry in adMixed Populations (LAMP) analysis was performed using hg19 genomic coordinates and strand alignment was completed with ShapeIT v2 (44, 45). Only SNPs common to both the 1000 Genomes Project reference panel and the FPC patient cohort were analyzed (669,977 SNPs; ref. 19). Ancestral allele frequencies were defined using 1000 Genomes project EUR, AFR, and ASN population groups. LAMP analysis was run using the following parameters: (i) three populations (EUR, AFR, and ASN); (ii) 10 generations of ancestral population mixing; (iii) African-American, Asian, and Caucasian proportions in FPC patient cohort of 0.028, 0.012, and 0.960 based on self-reported ancestry; (iv) recombination rate of 1×10^{-8}; and (v) LD cutoff of 0.1. Chromosomes were analyzed separately and then combined to obtain an average proportion of ancestry from each population for each cohort member.

Whole Exome Sequencing of FPC Patient Tumor Samples

Whole exome captured DNA libraries were prepared from nontumor tissue and microdissected fresh-frozen, formalin-fixed paraffin-embedded, or cell lines prepared from pancreatic adenocarcinomas from individuals with FPC. Cell line identity in relation to a primary patient sample was confirmed with Identifiler (cat. No. 4322288; Thermo Fisher Scientific) prior to sequencing. Library construction,

Downloaded from http://aacrjournals.org/cancerdiscovery/article-pdf/6/2/166/1824048/166.pdf by guest on 24 August 2022

sequencing, and bioinformatic analyses were performed at Personal Genome Diagnostics. In brief, genomic DNA from tumor and normal samples were fragmented and used for Illumina TruSeq library construction (Illumina). Exomic regions were captured in solution using the Agilent SureSelect v.4 kit according to the manufacturer's instructions (Agilent Technologies). Paired-end sequencing, resulting in 100 bp from each end of the fragments, was performed using a HiSeq 2000 Genome Analyzer (Illumina). Sequences were aligned to the human genome reference sequence (hg19) using the Eland algorithm of CASAVA 1.7 software (Illumina). The chastity filter of the BaseCall software of Illumina was used to select sequence reads for subsequent analysis. The ELANDv2 algorithm of CASAVA 1.7 software (Illumina) was then applied to identify point mutations and small insertions and deletions. Known polymorphisms recorded in the dbSNPs were removed from the analysis (17). Potential somatic mutations were filtered and visually inspected as described previously (33). Copy-number alterations were identified by comparing normalized average per-base coverage for a particular gene in a tumor sample to the normalized average per-base coverage in a matched normal sample for the patient.

Whole Exome Sequencing of BCCS Samples

As many as 967 individuals were selected from the BCCS, also known as the Rare BLISS sample, to act as controls (30). DNA libraries were prepared by processing genomic DNA samples into Illumina paired-end libraries using Illumina-compatible barcoded DNA adapters. Briefly, 1 to 3 μg purified genomic DNA was initially fragmented using a Covaris S2 instrument (Covaris), followed by end repair and ligation to paired-end adapters. Precapture libraries were enriched with an additional eight cycles of high-fidelity PCR, and quality and yield were assessed using the Bioanalyzer DNA 1000 Kit (catalog No. 5067-1504; Agilent Technologies) and the NanoDrop 1000 Spectrophotometer (Thermo Fisher Scientific). Exome capture was performed with the SeqCap EZ Human Exome Library v2.0 (catalog No. 05860504001; Roche Sequencing). Captured DNA libraries were sequenced with the Illumina HiSeq 2000 (Illumina), generating 152 bp (2 × 76 bp reads) per fragment in the final library. Sequence reads were processed through a standardized variant calling pipeline at either Cold Spring Harbor Laboratories or the Johns Hopkins University. Sequence reads were aligned to the human reference genome (UCSC hg19) using Burrows–Wheeler Aligner (BWA), allowing for two mismatches in the 30-base seed (46). Picard was used to correct mate pair mismatch, remove duplicate reads, and assess target region coverage (47). Samples with ≥75% of the target region covered at ≥20× sequencing were used for analyses. The Genome Analysis Toolkit (GATK) was used to generate SNVs and small INDEL calls in the target regions, after local realignment around INDELs and base score recalibration with the Unified Genotyper (48). The following GATK filters were used: variant confidence score ≥30, mapping quality ≥40, read depth ≥6, strand bias FSfilter <60. SNV clusters, defined as greater than three SNVs per 10 bases, and SNVs falling within a called INDEL region, were masked. Variants were annotated as described for whole genome sequencing of patients with FPC.

Confirmation of TET2 *and* ASXL1 *Variants*

Confirmation of variants was performed on DNA from blood and formalin-fixed paraffin embedded tissues using the Safe-Sequencing System (Safe-SeqS) as previously described (49). Primer sequences used to detect the *TET2* (g.chr4:106196537_C>T; p.Q1624X) variant were: cacacaggaaacagctatgaccatgGGGGAGAATAGGAACCCAGA and cgacgtaaaacgacggccagtNNNNNNNNNNNNNNAATCCCATGAACCCTTACCC. Primer sequences used to detect the *ASXL1* variant (g.chr20:31022414_T>TA; p.fs) were: cacacaggaaacagctatgaccatgCTCTGCCACCTCCCTCATC and cgacgtaaaacgacggccagtNNNNNNNNNNNNGGACCCTCGCAGACATTAAA. Ns denote degenerate bases, with an equal representation of A, C, T, and G.

Statistical Analyses

Two-sided *P* values were calculated using a Fisher exact test. False discovery rate was calculated using the Benjamini–Hochberg procedure. A *P* value of less than 0.05 was considered significant.

Data Availability

Whole genome and exome sequencing data are available (50). Users must obtain Institutional Review Board approval from their institutions and agree to policies that maintain patient privacy prior to use.

Disclosure of Potential Conflicts of Interest

W.R. McCombie has received honoraria from the speakers bureaus of Illumina and Pacific Biosciences and is a consultant/advisory board member for RainDance Technologies, Inc. and Orion Genomics. L.D. Wood is a consultant/advisory board member for Personal Genome Diagnostics (PGDx). M. Goggins is a consultant/advisory board member for Myriad Genetics. N. Papadopoulos has ownership interest (including patents) in PGDx and PapGene, Inc. and is a consultant/advisory board member for PGDx, PapGene, Inc, and Sysmex Inc. K.W. Kinzler has ownership interest in PGDx and in a PALB2 patent and is a consultant/advisory board member for PGDx. B. Vogelstein has ownership interest in Personal Genome Diagnostics, Inc. and in a PALB2 patent and is a consultant/advisory board member for Personal Genome Diagnostics, Inc. R.H. Hruban has ownership interest (including patents) in Myriad Genetics. A.P. Klein has ownership interest (including patents) in Myriad Genetics. No potential conflicts of interest were disclosed by the other authors.

Authors' Contributions

Conception and design: N.J. Roberts, G.M. Petersen, S. Gallinger, A.G. Schwartz, J.M. Herman, N. Papadopoulos, K.W. Kinzler, B. Vogelstein, R.H. Hruban, A.P. Klein

Development of methodology: N.J. Roberts, G.M. Petersen, J.M. Herman, J. Parla, N. Papadopoulos, A.P. Klein

Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): N.J. Roberts, A.L. Norris, G.M. Petersen, M.L. Bondy, R. Brand, S. Gallinger, R.C. Kurtz, S.H. Olson, A.K. Rustgi, A.G. Schwartz, E. Stoffel, S. Syngal, G. Zogopoulos, J. Axilbund, M.L. Cote, F.S. Goes, J.M. Herman, C. Iacobuzio-Donahue, A. Makohon-Moore, M. Pirooznia, J.B. Potash, A.D. Rhim, A.L. Smith, C.L. Wolfgang, L.D. Wood, P.P. Zandi, M. Goggins, J.R. Eshleman, N. Papadopoulos, A.P. Klein

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): N.J. Roberts, A.L. Norris, G.M. Petersen, S. Gallinger, S. Syngal, Y.-C. Chen, E.J. Childs, C. Douville, A. Makohon-Moore, K.W. McMahon, N. Niknafs, M. Pirooznia, Y. Wang, C.L. Wolfgang, M. Goggins, R. Karchin, J.R. Eshleman, N. Papadopoulos, K.W. Kinzler, R.H. Hruban, A.P. Klein

Writing, review, and/or revision of the manuscript: N.J. Roberts, G.M. Petersen, M.L. Bondy, R. Brand, S. Gallinger, S.H. Olson, A.K. Rustgi, A.G. Schwartz, E. Stoffel, S. Syngal, G. Zogopoulos, S.Z. Ali, J. Axilbund, K.G. Chaffee, M.L. Cote, E.J. Childs, C. Iacobuzio-Donahue, A. Makohon-Moore, M. Pirooznia, J.B. Potash, A.D. Rhim, C.L. Wolfgang, L.D. Wood, P.P. Zandi, M. Goggins, J.R. Eshleman, N. Papadopoulos, K.W. Kinzler, B. Vogelstein, R.H. Hruban, A.P. Klein

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): K.G. Chaffee, M. Pirooznia, K.W. Kinzler, A.P. Klein

Study supervision: S. Syngal, N. Papadopoulos, K.W. Kinzler, A.P. Klein

Other (was part of the project which generated exome data that were used for comparison in this paper): M. Kramer

Other (is co-PI on the BCCS that generated the whole exome sequencing data used for comparison of findings in this study): W. R. McCombie

Downloaded from http://aacrjournals.org/cancerdiscovery/article-pdf/6/2/166/1824048/166.pdf by guest on 24 August 2022

Acknowledgments

The authors thank all study participants for their generous contribution to this work. They also thank S. Angiuoli, C. Michael, M. Borges, L. Dobbyn, D. Echavarria, C. Harrington, S. Jones, M. Popoli, J. Ptak, R. Romans, J. Schaefer, and N. Silliman for technical assistance.

Grant Support

This work was generously supported by Dennis Troper and Susan Wojcicki, the Lustgarten Foundation for Pancreatic Cancer Research, the Sol Goldman Pancreatic Cancer Research Center, the Howard Hughes Medical Institute, the Virginia and D.K. Ludwig Fund for Cancer Research, the Stringer Foundation, the Rolfe Foundation for Cancer Research, the Joseph C. Monastra Foundation, the Gerald O. Mann Charitable Foundation (Harriet and Allan Wulfstat, Trustees), the Ladies Auxiliary to the Veterans of Foreign Wars, the friends and family of Roger L. Kerns Sr., the Weston Garfield Foundation, the NIH Specialized Programs of Research Excellence P50-CA062924 and P50-CA102701, and NIH grants K99-CA190889, K01-MH093809, P30-CA006973, R01-CA57345, R01-CA97075, R01-CA154823, R01-DK060694, and R01-MH087979.

Received April 3, 2015; revised November 25, 2015; accepted December 2, 2015; published OnlineFirst December 9, 2015.

REFERENCES

1. Cancer facts and figures 2015. 2015; 1–56. Available from: http://www.cancer.org/research/cancerfactsstatistics/cancerfactsfigures2015/index/
2. Hruban RH, Zamboni G. Pancreatic cancer. Special issue–insights and controversies in pancreatic pathology. Arch Pathol Lab Med 2009;133:347–9.
3. Klein AP, Brune KA, Petersen GM, Goggins M, Tersmette AC, Offerhaus GJ, et al. Prospective risk of pancreatic cancer in familial pancreatic cancer kindreds. Cancer Res 2004;64:2634–8.
4. Yuen RK, Thiruvahindrapuram B, Merico D, Walker S, Tammimies K, Hoang N, et al. Whole-genome sequencing of quartet families with autism spectrum disorder. Nat Med 2015;21:185–91.
5. Canto MI, Harinck F, Hruban RH, Offerhaus GJ, Poley JW, Kamel I, et al. International Cancer of the Pancreas Screening (CAPS) Consortium summit on the management of patients with increased risk for familial pancreatic cancer. Gut 2013;62:339–47.
6. Wang L, Brune KA, Visvanathan K, Laheru D, Herman J, Wolfgang C, et al. Elevated cancer mortality in the relatives of patients with pancreatic cancer. Cancer Epidemiol Biomarkers Prev 2009;18:2829–34.
7. Waddell N, Pajic M, Patch AM, Chang DK, Kassahn KS, Bailey P, et al. Whole genomes redefine the mutational landscape of pancreatic cancer. Nature 2015;518:495–501.
8. Jones S, Hruban RH, Kamiyama M, Borges M, Zhang X, Parsons DW, et al. Exomic sequencing identifies PALB2 as a pancreatic cancer susceptibility gene. Science 2009;324:217.
9. Roberts NJ, Jiao Y, Yu J, Kopelovich L, Petersen GM, Bondy ML, et al. ATM mutations in patients with hereditary pancreatic cancer. Cancer Discov 2012;2:41–6.
10. Murphy KM, Brune KA, Griffin C, Sollenberger JE, Petersen GM, Bansal R, et al. Evaluation of candidate genes MAP2K4, MADH4, ACVR1B, and BRCA2 in familial pancreatic cancer: deleterious BRCA2 mutations in 17%. Cancer Res 2002;62:3789–93.
11. Al-Sukhni W, Rothenmund H, Borgida AE, Zogopoulos G, O'Shea AM, Pollett A, et al. Germline BRCA1 mutations predispose to pancreatic adenocarcinoma. Hum Genet 2008;124:271–8.
12. McWilliams RR, Wieben ED, Rabe KG, Pedersen KS, Wu Y, Sicotte H, et al. Prevalence of CDKN2A mutations in pancreatic cancer patients: implications for genetic counseling. Eur J Hum Genet 2011;19:472–8.
13. Kastrinos F, Mukherjee B, Tayob N, Wang F, Sparr J, Raymond VM, et al. Risk of pancreatic cancer in families with Lynch syndrome. JAMA 2009;302:1790–5.
14. Giardiello FM, Brensinger JD, Tersmette AC, Goodman SN, Petersen GM, Booker SV, et al. Very high risk of cancer in familial Peutz-Jeghers syndrome. Gastroenterology 2000;119:1447–53.
15. Ruijs MW, Verhoef S, Rookus MA, Pruntel R, van der Hout AH, Hogervorst FB, et al. TP53 germline mutation testing in 180 families suspected of Li–Fraumeni syndrome: mutation detection rate and relative frequency of cancers in different familial phenotypes. J Med Genet 2010;47:421–8.
16. Lowenfels AB, Maisonneuve P, Cavallini G, Ammann RW, Lankisch PG, Andersen JR, et al. Pancreatitis and the risk of pancreatic cancer. International Pancreatitis Study Group. N Engl J Med 1993;328:1433–7.
17. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, et al. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res 2001;29:308–11.
18. Rahman N. Realizing the promise of cancer predisposition genes. Nature 2014;505:302–8.
19. The 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. Nature 2012;491:56–65.
20. NHLBI Exome Sequencing Project (ESP) [Internet]. Seattle: University of Washington[cited 2015 Nov 18]. Available from: http://evs.gs.washington.edu/EVS/
21. Derrien T, Estelle J, Marco Sola S, Knowles DG, Raineri E, Guigo R, et al. Fast computation and applications of genome mappability. PLoS One 2012;7:e30377.
22. Kent WJ. BLAT–the BLAST-like alignment tool. Genome Res 2002;12:656–64.
23. Ruark E, Snape K, Humburg P, Loveday C, Bajrami I, Brough R, et al. Mosaic PPM1D mutations are associated with predisposition to breast and ovarian cancer. Nature 2013;493:406–10.
24. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA, Kinzler KW. Cancer genome landscapes. Science 2013;339:1546–58.
25. Shlush LI, Zandi S, Mitchell A, Chen WC, Brandwein JM, Gupta V, et al. Identification of pre-leukaemic haematopoietic stem cells in acute leukaemia. Nature 2014;506:328–33.
26. Busque L, Patel JP, Figueroa ME, Vasanthakumar A, Provost S, Hamilou Z, et al. Recurrent somatic TET2 mutations in normal elderly individuals with clonal hematopoiesis. Nat Genet 2012;44:1179–81.
27. Xie M, Lu C, Wang J, McLellan MD, Johnson KJ, Wendl MC, et al. Age-related mutations associated with clonal hematopoietic expansion and malignancies. Nat Med 2014;20:1472–8.
28. Jaiswal S, Fontanillas P, Flannick J, Manning A, Grauman PV, Mar BG, et al. Age-related clonal hematopoiesis associated with adverse outcomes. N Engl J Med 2014;371:2488–98.
29. Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM, et al. ClinVar: public archive of relationships among sequence variation and human phenotype. Nucleic Acids Res 2014;42:D980–5.
30. Chen YC, Carter H, Parla J, Kramer M, Goes FS, Pirooznia M, et al. A hybrid likelihood model for sequence-based disease association studies. PLoS Genet 2013;9:e1003224.
31. Knudson AG Jr. Mutation and cancer: statistical study of retinoblastoma. Proc Natl Acad Sci U S A 1971;68:820–3.
32. Norris AL, Roberts NJ, Jones S, Wheelan SJ, Papadopoulos N, Vogelstein B, et al. Familial and sporadic pancreatic cancer share the same molecular pathogenesis. Fam Cancer 2015;14:95–103.
33. Jones S, Zhang X, Parsons DW, Lin JC, Leary RJ, Angenendt P, et al. Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. Science 2008;321:1801–6.
34. Biankin AV, Waddell N, Kassahn KS, Gingras MC, Muthuswamy LB, Johns AL, et al. Pancreatic cancer genomes reveal aberrations in axon guidance pathway genes. Nature 2012;491:399–405.
35. Rio Frio T, Lavoie J, Hamel N, Geyer FC, Kushner YB, Novak DJ, et al. Homozygous BUB1B mutation and susceptibility to gastrointestinal neoplasia. N Engl J Med 2010;363:2628–37.
36. Witt H, Beer S, Rosendahl J, Chen JM, Chandak GR, Masamune A, et al. Variants in CPA1 are strongly associated with early onset chronic pancreatitis. Nat Genet 2013;45:1216–20.
37. Duell EJ, Lucenteforte E, Olson SH, Bracci PM, Li D, Risch HA, et al. Pancreatitis and pancreatic cancer risk: a pooled analysis in the

Downloaded from http://aacrjournals.org/cancerdiscovery/article-pdf/6/2/166/1824048/166.pdf by guest on 24 August 2022

International Pancreatic Cancer Case–Control Consortium (PanC4). Ann Oncol 2012;23:2964–70.
38. Woodage T, King SM, Wacholder S, Hartge P, Struewing JP, McAdams M, et al. The APCI1307K allele and cancer risk in a community-based study of Ashkenazi Jews. Nat Genet 1998;20:62–5.
39. Giardiello FM, Offerhaus GJ, Lee DH, Krush AJ, Tersmette AC, Booker SV, et al. Increased risk of thyroid and pancreatic carcinoma in familial adenomatous polyposis. Gut 1993;34:1394–6.
40. Hruban RH, Canto MI, Goggins M, Schulick R, Klein AP. Update on familial pancreatic cancer. Adv Surg 2010;44:293–311.
41. Srivastava S, Zou ZQ, Pirollo K, Blattner W, Chang EH. Germ-line transmission of a mutated p53 gene in a cancer-prone family with Li-Fraumeni syndrome. Nature 1990;348:747–9.
42. Pruitt KD, Brown GR, Hiatt SM, Thibaud-Nissen F, Astashyn A, Ermolaeva O, et al. RefSeq: an update on mammalian reference sequences. Nucleic Acids Res 2014;42:D756–63.
43. Douville C, Carter H, Kim R, Niknafs N, Diekhans M, Stenson PD, et al. CRAVAT: cancer-related analysis of variants toolkit. Bioinformatics 2013;29:647–8.
44. Pasaniuc B, Sankararaman S, Kimmel G, Halperin E. Inference of locus-specific ancestry in closely related populations. Bioinformatics 2009;25:i213–21.
45. Delaneau O, Marchini J, Zagury JF. A linear complexity phasing method for thousands of genomes. Nat Methods 2012;9:179–81.
46. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 2010;26:589–95.
47. Picard [Internet]. Cambridge, MA: The Broad Institute; c2015[cited 2015 Nov 18]. Available from: http://broadinstitute.github.io/picard/
48. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 2010;20:1297–303.
49. Kinde I, Bettegowda C, Wang Y, Wu J, Agrawal N, Shih Ie M, et al. Evaluation of DNA from the Papanicolaou test to detect ovarian and endometrial cancers. Sci Transl Med 2013;5:167ra4.
50. The familial pancreatic cancer genome sequencing project [Internet]. Baltimore, MD: The Johns Hopkins University; [cited 2015 Nov 25]. Available from: http://www.familialpancreaticcancer.org/

Downloaded from http://aacrjournals.org/cancerdiscovery/article-pdf/6/2/166/1824048/166.pdf by guest on 24 August 2022